Path Integral Monte Carlo Methods for Fermions

D. M. Ceperley

National Center for Supercomputer Applications
Dept. of Physics, University of Illinois at Urbana-Champaign
1110 W. Green. St., Urbana, Il 61801

This article discusses the basic properties of the path integral method for continuum fermions, focusing on the restricted path integral (RPIMC) approach. The RPIMC is a simulation technique for fermion systems that extends the previously developed bosonic methods to degenerate many-fermion systems, avoiding the fermion sign problem by restricting the paths. For general fermion systems, the restriction needed for exact results is not known, so ansatz for the nodes of the fermion density matrix are used. Even with this approximation, the finite temperature Monte Carlo method is expected to be even more accurate and powerful than corresponding zero-temperature Monte Carlo methods.

CONTENTS

I. INTRODUCTION

The lectures at this school have concerned various types of simulations. We have heard about various attempts to perform quantum dynamical calculations, but what I want to tell you about are recent attempts to simulate the equilibrium properties of fermion systems. This remains one of the most challenging problems in the field and one of the most important both from a theoretical and practical point of view, given the pervasiveness of fermions in nature and intractability of research on many-fermion systems.

Much progress has been made in recent years in performing simulations of quantum many-body boson and distinguishable-particle systems using imaginary-time path integrals. As first shown by Feynman (1953), and as we will later describe, the many-body density matrix can be obtained using classical-statistical methods on polymer-like systems. The density matrix of a many-body system at a temperature $k_B T = \beta^{-1}$ can be written as an integral over all paths $\{R_t\}$:

$$\rho(R_0, R_\beta; \beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \oint_{\mathcal{P}R_0 \to R_\beta} dR_t \exp(-S[R_t]). \tag{1.1}$$

The path $R(t)$ begins at $\mathcal{P}R_0$ and ends at R_β, and $\mathcal{P}$ is a permutation of particle labels. For N particles, the path is in $3N$ dimensional space: $R_t = (\mathbf{r}_{1t}, \mathbf{r}_{2t} \ldots \mathbf{r}_{Nt})$. The upper sign is to be used for bosons and the lower sign for fermions. For nonrelativistic particles interacting with a potential $V(R)$, the ***action*** of the path, $S[R_t]$, is given by:

$$S[R_t] = \int_0^\beta dt \left[\frac{m}{2} \left| \frac{dR(t)}{\hbar dt} \right|^2 + V(R_t) \right]. \tag{1.2}$$

Thermodynamic properties, such as the energy, are related to the diagonal part of the density matrix, so that the path returns to its starting place or to a permutation $\mathcal{P}$ of its starting place after a "time" β.

Since the imaginary-time ***action*** $S[R_t]$ is a real function of the path, for boltzmannons and bosons the integrand is nonnegative. It can thus be interpreted as a probability of an equivalent classical system and the action as the classical potential energy of a "polymer." To perform Monte Carlo calculations of the integrand, one makes imaginary time discrete, so that one has a finite (and hopefully small) number of time slices and thus a classical system of N particles in M time slices; an equivalent NM particle classical system of "polymers." If the path integral is performed by a simulation method, such as a generalization of Metropolis Monte Carlo or with Molecular Dynamics, one can obtain essentially exact results for previously unsolved problems such as the properties of liquid ^{4}He at temperatures near the superfluid phase transition, the exchange frequency in quantum crystals, and quantum particles immersed in classical systems. (Ceperley, 1995)

Note that in addition to sampling the path, the permutation is also sampled. This is equivalent to allowing the ring polymers to connect in different ways. This macroscopic "percolation" of the polymers is directly related to superfluidity as Feynman (1953) first showed. Any permutation can be broken into permutation cycles, *i. e.* into 2-, 3-, ... exchange cycles. Superfluid behavior can occur at low temperature when the probability of exchange cycles on the order of the system size is non-negligible. Figure 1 shows a trace of a typical path of six "superfluid" helium atoms in a 2D periodic square. I have recently written a detailed article describing the path integral theory of Bose superfluids and how one carries out the Monte Carlo calculations (Ceperley, 1995).

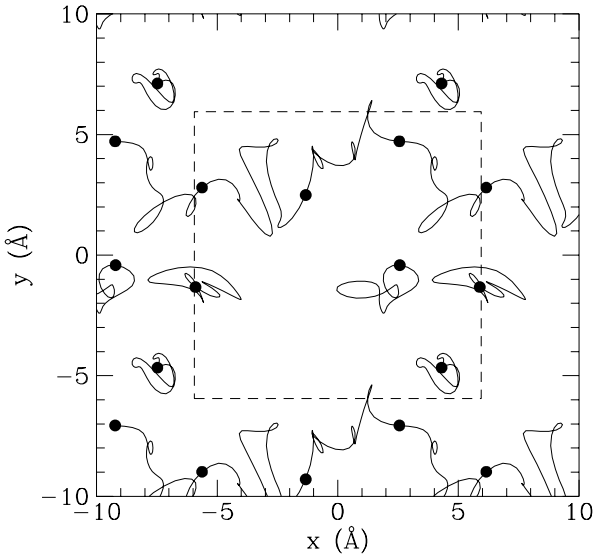

FIG. 1. The extended trace of six ^{4}He atoms at a temperature of 0.75 K. The dashed square represents the periodic boundary conditions. The six filled circles represent the beginning of a path for each atom. Three of the atoms are involved in an exchange which winds around the boundary in the x direction. The paths have been Fourier smoothed for clarity.

However, the straightforward application of those techniques to Fermi systems means that odd permutations subtract from the integrand. This is the "fermion sign problem" which I will derive in some detail later on. Path integral methods as rigorous and successful as those for boson systems are not yet known for fermion systems in spite of the activities of many scientists throughout the last four decades. On the other hand there has been substantial development of zero temperature fermion simulations using the fixed-node method(Anderson,1976). A well-known example is the calculation of the correlation energy of the uniform electron gas using Diffusion Monte Carlo (Ceperley and Alder, 1980). In the fixed-node approach, one solves the Schroedinger equation with the boundary condition that the many-body wavefunction vanish when a trial wavefunction does. This gives the best upper bound consistent with the assumed nodes. With the nodal assumption, the fixed-node approach does not have the fermion sign. Since there are several recent reviews on ground state methods (Schmidt and Kalos,1987; Senatore and March, 1994; Hammond, Lester, and Reynolds, 1994; Anderson, 1995), I will not discuss those here.

Path Integrals have some significant advantages over zero temperature methods (as well as disadvantages, of course). Among the advantages is the absence of a trial wavefunction which means that quantum expectation values, including ones not involving the energy, can be directly computed. For the expert, the lack of an importance function may seem a disadvantage; without it one cannot push the simulation in a preferred direction. However as the quantum system becomes more complex, it becomes increasingly difficult to devise satisfactory trial functions. It is my opinion that if quantum simulations are to attack the type of complex physical situations in which classical simulations routinely deal with, it is better to have a formulation without a trial function. Only the Hamiltonian should enter. Of course, the explicit formulation at finite temperature also makes comparison with experiment more direct.

I will briefly mention various ways around the fermion sign problem, but concentrate on the restricted fermion path integral method. This approach has not been described in much detail previously, so I will take the opportunity to do so here. In addition to giving some simple examples of the application of restricted paths, I will discuss a few practical aspects of fermion path integrals such as sampling and nodal actions. Finally, I will briefly review some of the applications in order to point out the advantages of RPIMC, the future directions, and its problems. I shall only discuss continuum fermion systems in equilibrium and not discuss lattice models or dynamical aspects. Gubernatis and deRaedt cover those topics elsewhere in this volume.

II. IMAGINARY TIME PATH INTEGRALS

In this lecture, we shall consider a system of N nonrelativistic particles interacting with a potential $V(R)$. The Hamiltonian is:

$$\mathcal{H} = -\sum_{i=1}^{N} \lambda_i \nabla_i^2 + V(R) \tag{2.1}$$

where we shall use the definition $\lambda_i = \hbar^2/2m_i$.

Thermodynamic properties are averages over the thermal N-body density matrix which is defined as a thermal occupation of the exact eigenstates $\phi_i(R)$:

$$\rho(R, R'; \beta) = \sum_i \phi_i^*(R) e^{-\beta E_i} \phi_i(R'). \tag{2.2}$$

The partition function is the trace of the density matrix.

$$Z(\beta) = e^{-\beta F} = \int dR \rho(R, R; \beta) = \sum_i e^{-\beta E_i} \tag{2.3}$$

Other thermodynamic averages are obtained as:

$$\langle \mathcal{O} \rangle = Z(\beta)^{-1} \int dR dR' \langle R|\mathcal{O}|R'\rangle \rho(R', R; \beta). \tag{2.4}$$

As mentioned above, the density matrix can be calculated using path integrals.

Now let us consider how particle statistics are expressed in path integrals. For systems of identical particles, the states can be classified into symmetric and antisymmetric states. The fermion density matrix is defined by restricting the sum to be only over antisymmetric states. (Similarly for other symmetries such as momentum or spin.) We shall denote the statistics of the particles by subscripts: ρ_F will denote the fermion density matrix, ρ_B the boson density matrix, ρ_D the boltzmannon (distinguishable particle) density matrix, and ρ any of the above density matrices.

Note that for any density matrix the diagonal part is always positive:

$$\rho(R, R; \beta) \geq 0 \tag{2.5}$$

so that $Z^{-1}\rho(R, R; \beta)$ is a proper probability distribution. It is the diagonal part which we need for many observables, so that probabilistic ways of calculating those observables are, in principle, possible.

For the moment, let us consider a single component system of spinless fermions. (By spinless fermions, we mean that the spatial wavefunction is antisymmetric with respect to interchange of all pairs of spatial variables.) Let $\mathcal{P}$ be one of the $N!$ permutations of particle labels. Then each of the fermion eigenstates has the following property:

$$\phi(\mathcal{P}R) = (-1)^{\mathcal{P}} \phi(R). \tag{2.6}$$

Using Eqs. (2.2) and (2.6) the density matrix has the following symmetries:

$$\begin{aligned} \rho(R, R'; \beta) &= \rho(R', R; \beta) \\ \rho_F(R, R'; \beta) &= (-1)^{\mathcal{P}} \rho_F(\mathcal{P}R, R'; \beta) \\ \rho_F(R, R'; \beta) &= (-1)^{\mathcal{P}} \rho_F(R, \mathcal{P}R'; \beta). \end{aligned} \tag{2.7}$$

One can use the permutation (or relabeling) operator to construct the path integral expression for the boson or fermion density matrix in terms of the Boltzmann density matrix:

$$\rho_{B/F}(R, R'; \beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \rho_D(\mathcal{P}R, R'; \beta). \tag{2.8}$$

More generally, one uses some projection operator to select a desired set of states from the distinguishable particle density matrix which contains all states. From Eq. (2.7) we could (anti)symmetrize with respect to the first argument, the last argument or both. This connection between the boltzmannon density matrix and the Bose/Fermion density matrix is important because it is the boltzmannon density matrix that is built naturally from paths. In this lecture, except for how paths close, particles are generally considered to be distinguishable. This is in contrast to the second-quantized philosophy, where one always works with an antisymmetric basis. The distinguishable basis has advantages for numerical computation.

An alternative definition of the density matrix is by its evolution in imaginary time, the Bloch equation:

$$-\frac{\partial \rho(R, R'; t)}{\partial t} = \mathcal{H} \rho(R, R'; t) \tag{2.9}$$

which obeys the boundary condition at $t = 0$ for boltzmannon statistics:

$$\rho_D(R, R'; 0) = \delta(R - R') \tag{2.10}$$

or for Bose or Fermi statistics:

$$\rho_{B/F}(R, R'; 0) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \delta(\mathcal{P}R - R'). \tag{2.11}$$

The high temperature boundary condition is an (anti)symmetrized delta function.

Path integrals are constructed using the product property of density matrices:

$$\rho(R_0, R_2, \beta_1 + \beta_2) = \int dR_1 \rho(R_0, R_1; \beta_1) \rho(R_1, R_2; \beta_2). \tag{2.12}$$

The product property holds for any sort of density matrix.

If the product property is used M times we can relate the density matrix at a temperature β^{-1} to the density matrix at $M\beta^{-1}$:

$$\begin{aligned} \rho_{B/F}(R_0, R_M, \beta) = \frac{1}{N!} \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \int dR_1 \ldots dR_{M-1} \\ \rho_D(\mathcal{P}R_0, R_1; \tau) \ldots \rho_D(R_{M-1}, R_M; \tau). \end{aligned} \tag{2.13}$$

The sequence of intermediate points $\{R_1, R_2, \ldots, R_{M-1}\}$ is the path, and the ***time step*** is $\tau = \beta/M$.

As the time step gets sufficiently small, we can write down an explicit expression for the density matrix ρ_D and thereby an explicit expression for $\rho(R_0, R_M; \beta)$, but one with lots of intermediate integrals and the permutational sum to perform. The Trotter theorem tells us that for sufficiently small τ we can assume that the kinetic and potential operators commute so that: $e^{-\tau\mathcal{H}} = e^{\tau\mathcal{T}} e^{-\tau\mathcal{V}}$. Define the (boltzmannon) action as $S_D(R, R'; \tau) = -\ln[\rho_D(R, R'; \tau)]$. Then for small time step the action is:

$$S_D(R, R'; \tau) = \frac{3N}{2} \ln(4\pi\lambda\tau) + \frac{(R - R')^2}{4\lambda\tau} + \frac{1}{2}(V(R) + V(R')). \tag{2.14}$$

This is known as the ***primitive approximation*** to the action. The form of the action is analogous to the potential energy of classical "polymer" system with harmonic springs between neighboring beads and an interpolymer potential between different chains.

The boson action is real, but it is expressed as a sum over permutations. For large N it is not possible to evaluate direcly the sum since it has $N!$ terms. It is better to leave the bosonic symmetrization as an explicit boundary condition on the paths and to sample the permutations as well as the paths. We will follow the same philosophy with restricted fermion paths, the reasons not being the difficulty of evaluating the resulting determinant (that is easier for fermions than for bosons) but to avoid the minus signs. For bosons or fermions we can (anti)symmetrize anywhere along the path as many times as we like. However relabeling is only necessary once. By convention we will just relabel the first step.

The Feynman-Kacs formula for the density matrix results from taking the limit $M \to \infty$. The action for the path can be interpreted as as Brownian (diffusion) motion weighted by potential energy along the path:

$$\rho_D(R_0, R_\beta; \beta) = \rho_D^0(R_0, R_\beta; \beta) \left\langle \exp[-\int_0^\beta dt V(R_t)] \right\rangle_{BW}. \tag{2.15}$$

where the brackets mean to average over all Brownian (Gaussian) walks and ρ_D^0 is the free particle density matrix (*i.e.* $V = 0$).

III. THE DIRECT FERMION PATH INTEGRAL METHOD

In the ***direct fermion method*** one sums over permutations just as for bosonic systems. Odd permutations then contribute with a negative weight. The direct method has a major problem because of the cancellation of positive and negative permutations. This was first noted by Feynman and Hibbs (1965), page 292-293 who, after describing the path integral theory for boson superfluid ^{4}He, noted:

> The (path integral) expression for Fermi particles, such as He^3, is also easily written down. However, in the case of liquid He^3, the effect of the potential is very hard to evaluate quantitatively in an accurate manner. The reason for this is that the contribution of a cycle to the sum over permutations is either positive or negative depending on whether the cycle has an odd or even number of atoms in its length L. At very low temperature, the contributions of cycles such as $L = 51$ and $L = 52$ are very nearly equal but opposite in sign, and therefore they very nearly cancel. It is necessary to compute the difference between such terms, and this requires very careful calculation of each term separately. It is very difficult to sum an alternating series of large terms which are decreasing slowly in magnitude when a precise analytic formula for each term is not available.
>
> Progress could be made in this problem if it were possible to arrange the mathematics describing a Fermi system in a way that corresponds to a sum of positive terms. Some such schemes have been tried, but the resulting terms appear to be much too hard to evaluate even qualitatively.
>
> . . .
>
> The (explanation) of the superconducting state was first answered in a convincing way by Bardeen, Cooper, and Schrieffer. The path integral approach played no part in their analysis, and ***in fact has never proved useful for degenerate Fermi systems.*** [my italics]

As Feynman argued above, for the calculation of any operator by the direct fermion Path Integral Monte Carlo there is a tremendous loss of efficiency over the boson case. Here I demonstrate that a distribution having both positive and negative regions will have an exponentially vanishing (in a sense to be specified) signal/noise ratio in a Monte Carlo calculation.

First, we need to show that the optimal sampling function is the absolute value of the integrand which for fermion path integrals is just the the bosonic path probability distribution. Consider the evaluation of an expectation value for an operator $\mathcal{O}$:

$$\langle \mathcal{O} \rangle = \frac{\int dx \pi \mathcal{O}}{\int dx \pi} \tag{3.1}$$

where π is a function with both positive and negative pieces and $\int dx$ indicates not only an integral over coordinates, but also the sum over the permutations. Now

consider doing importance sampling with some arbitrary sampling function $P(x)$. By sampling function, we mean that $P(x) \geq 0$ and $\int dx P(x) = 1$. The estimator for $\mathcal{O}$ can be written as:

$$\langle \mathcal{O} \rangle = \lim_{M \to \infty} \frac{\sum_{i=1}^{M} w(x_i) \mathcal{O}(x_i)}{\sum_{i=1}^{M} w(x_i)} \tag{3.2}$$

where the weight is: $w(x_i) = \pi(x_i)/P(x_i)$ and $\{x_i\}$ is a set of M points sampled from $P(x)$. The variance of the estimator is:

$$\nu_{\mathcal{O}} = \frac{1}{M[\int \pi]^2} \int dx \frac{[\pi(x)(\mathcal{O}(x) - \langle \mathcal{O} \rangle)]^2}{P(x)} \tag{3.3}$$

For simplicity, we have assumed the sample points are statistically uncorrelated with each other. Otherwise we would have to include an additional correlation factor.

Let us now vary the distribution $P(x)$ to minimize the error. The sampling function minimizing the variance is found to be proportional to:

$$P^*_{\mathcal{O}}(x) \propto |\pi(x)(\mathcal{O}(x) - \langle \mathcal{O} \rangle)|. \tag{3.4}$$

The optimal distribution depends on the desired operator $\mathcal{O}$. Normally, one wants a sampling function which is good for a variety of properties and does not depend on the details of a single operator. A physical operator $\mathcal{O}$, is usually more smoothly varying than $\pi(x)$, which is sharply peaked in path space. Also, $\langle \mathcal{O} \rangle$ is not precisely known so it would have to be self-consistently determined by the sampling. For these reasons, one should ignore the second factor and choose $P(x) \propto |\pi(x)|$. With this choice the weights are $w(x) = sign(\pi(x)) = \pm 1$. Then substituting into the variance equation we obtain:

$$\nu_{\mathcal{O}} = \frac{\int dx |\pi(x)|}{M[\int dx \pi(x)]^2} \int dx |\pi(x)| (\mathcal{O}(x) - \langle \mathcal{O} \rangle)^2. \tag{3.5}$$

Now for a bosonic system: $\pi(x) \geq 0$ so we will write the bosonic variance as:

$$\nu^B_{\mathcal{O}} = \frac{1}{M \int |\pi|} \int |\pi(x)| (\mathcal{O}(x) - \langle \mathcal{O} \rangle)^2. \tag{3.6}$$

Then we can write the fermion variance as:

$$\nu_{\mathcal{O}} = \nu^B_{\mathcal{O}} / \xi \tag{3.7}$$

where the ***efficiency*** (related to the sign of π) is

$$\xi = \left[\frac{\int \pi}{\int |\pi|} \right]^2 = \left[\frac{M_+ - M_-}{M} \right]^2. \tag{3.8}$$

M_+/M is the average time the simulation spends in the positive region of π and M_-/M the average time in the negative region. The fermion efficiency is proportional to the square of the average sign: the excess of positive sampled points over negative sampled points.

Now let us specialize to the direct fermion method in Eq. (2.14). The sign of π is equal to the sign of the permutation, so the efficiency is simply the number of even permutations minus number of odd permutations. In fact, we can say the efficiency is exactly equal to:

$$\xi = \left[\frac{Z_F}{Z_B} \right]^2 = \exp[-2\beta(F_F - F_B)] \tag{3.9}$$

where Z_F and Z_B refer to partition function and F_F and F_B to the total free energies for Fermi and Bose statistics respectively. Of course the free energies are proportional to the number of particles. At high temperatures, it can be shown (Landau and Lifshitz, 1970, sec. 55) that:

$$\xi = \exp[-2\rho N (2\pi\lambda\beta)^{3/2}]. \tag{3.10}$$

At the degeneracy temperature: $\xi \approx \exp(-N)$! Loh et al (1990) and Newman and Kuki(1992) show figures of the average sign for lattice models and for two fermion systems. The direct fermion method, while exact, becomes exceedingly inefficient as β and N increase, precisely when the physics becomes interesting.

Sampling the absolute value of an integrand is the best we can do without using particular properties of the operator $\mathcal{O}$. Correlation between the points of the random walk used to sample π will only further reduce the efficiency. (In the fermion case one might do a bit better by enhancing the probability for tunneling between the positive and negative regions of the integrand.) The above estimate is the best case of how the variance depends on computer time, temperature, and the number of particles. One can see a loss of efficiency just due to the change of sign of π. Thus, with direct sampling, simulations of many-fermion systems at low temperatures are essentially hopeless because the computational efficiency goes to zero too rapidly.

IV. THE RESTRICTED PATH INTEGRAL METHOD

We now derive the restricted path identity: that the nodes of the exact density matrix determine the rule by which one can take only paths with the same sign. For the diagonal density matrix we can arrange things so that we only get positive contributions. We will do the proof for any symmetry, not just antisymmetry, although this lecture is only concerned with fermions.

Suppose ρ_F is the density matrix corresponding to some set of quantum numbers which is obtained by using the projection operator $\mathcal{A}$ on the distinguishable particle density matrix. Then it is a solution to the Bloch equation (2.9) with the initial condition:

$$\rho_F(R, R_*; 0) = \mathcal{A}\delta(R - R_*). \tag{4.1}$$

The point R_*, the right leg of the density matrix, known here as the ***reference point***, will be held fixed. Since $\mathcal{A}$ is a projection operator we have: $\rho_F(R_*, R_*; \beta) \geq 0$.

The argument that one can restrict that path integral is relatively simple, but for convenience we break it into several steps.

1. The solution to the Bloch equation (2.9) is uniquely determined by its boundary conditions. Consider some region Υ of "space-time". Then the solution in the interior is uniquely determined by the values on the boundaries of Υ. More than this, to determine the value of the density matrix at a space-time point (R, β), all that are needed are the boundary values for $\beta' < \beta$.

   A thought experiment with a computer big enough to store a fine grid in $3N$-dimensional space gives a intuitive proof of this result. Suppose at one time slice we know all values of $\rho_F(R)$ inside the spatial domain. Then we can use the Bloch equation and the boundary conditions to advance to the next time step. Hence the present plus the boundary conditions determine the future. See Ceperley (1991) for the proof of the uniqueness property.

2. The nodes of $\rho_F(R, R_*; t)$ carve up space-time into a finite number of nodal cells, $\Upsilon_k(R_*)$. Let us define a ***node-avoiding path***, as a continuous path R_t for $0 < t \leq \beta$ for which $\rho_F(R_t, R_*; t) \neq 0$ for all $0 < t < \beta$. We say that two points are in the same nodal cell if they are connected by a node-avoiding path. The collection of all space time points (R, t) connected by some node-avoiding path make up a nodal cell $\Upsilon_k(R_*)$. The surfaces of $\Upsilon_k(R_*)$ consist of points such that $\rho_F(R, R_*; t) = 0$.

   From i), we can solve the Bloch equation inside each nodal cell separately by specifying the initial conditions at $t = 0$ from Eq. (4.1) and zero boundary conditions on the surface of the nodal cell. We refer to the union of all nodal cells as $\Upsilon(R_*)$.

3. Now let us reconstruct the path integral expression for the density matrix taking into account the zero boundary conditions. To enforce the zero boundary condition, we can insert an infinite repulsive potential precisely at the nodal surfaces. But that will eliminate the contribution of any walks which hit or cross the node. Hence, demanding that a density matrix vanish on a surface is equivalent to restricting the path integral to be only over node-avoiding walks.

Thus we have proved the Restricted Path Integral identity:

$$\rho_F(R_\beta, R_*; \beta) = \int dR_0 \rho_F(R_0, R_*; 0) \oint_{R_0 \to R_\beta \in \Upsilon(R_*)} dR_t e^{-S[R_t]} \qquad (4.2)$$

where the subscript means that we restrict the path integration to paths starting at R_0, ending at R_β and are node-avoiding (those for which $\rho_F(R_t, R_*; t) \neq 0$ for all $0 < t \leq \beta$.) The weight of the walk is $\rho_F(R_0, R_*; 0)$. It is clear that the contribution of all the paths for a single element of the density matrix will be of the same sign; positive if $\rho_F(R_0, R_*; 0) > 0$, negative otherwise. This is what we set out to show. In particular, on the diagonal all contributions must be positive.

Important in this argument is that the random walk is a continuous process (the trajectory is continuous) so we can say definitively that if sign of the density matrix changed, it had to have crossed the node at some point. This presents a technical problem for discrete time paths since we must decide whether the path crosses and crosses back in between the sampled times. We take up this issue later. Lattice models or non-local Hamiltonians do not have continuous trajectories so this method is not as straightforward for those models.

The restricted path identity is one solution to Feynman's task of rearranging terms to keep only positive contributing paths for diagonal expectation values. The "bosonic" path integral formulation can be applied to fermion path integrals; all that is required is to take a subset of the bosonic paths. In principle, there exists a way to solve the "sign problem"! We shall see that it is important to allow non-trivial, even permutations. Macroscopic even permutations are directly related to Fermi liquid behavior.

The problem we now face is that the unknown density matrix appears both on the left-hand side and on the right-hand side of Eq. (4.2) since it is used to define the criterion of node-avoiding paths. To apply the formula directly, we would somehow have to self-consistently determine the density matrix. In practice what we need to do is make an ***ansatz***, which we call ρ_T, for the ***nodes*** of the density matrix needed for the restriction. The trial density matrix is used to define trial nodal cells: $\Upsilon_T(R_*)$. Using a trial nodes we generate a better approximation to the density matrix using Eq. (4.2) with the trial restriction:

$$\tilde{\rho}_T(R_\beta, R_*; \beta) = \int dR_0 \rho_F(R_0, R_*; 0) \oint_{R_0 \to R_\beta \in \Upsilon_T(R_*)} dR_t e^{-S[R_t]}. \qquad (4.3)$$

Hence, $\tilde{\rho}_T(R', R; \beta)$ is a solution to the Bloch equation inside the trial nodal cells, and it obeys the correct initial conditions. It is not an exact solution to the Bloch equation (unless the nodes of ρ_T are correct) because it has possible gradient discontinuities at the trial nodal surfaces.

By multiplying together $\tilde{\rho}_T$ and ρ_F, integrating over time and a common spatial variable, and using their respective definitions, one can show that $\tilde{\rho}_T$ is related to the exact solution by the following integral equation:

$$\tilde{\rho}_T(R_1,R_2;t) = \rho_F(R_1,R_2;t) + \int_0^t dt' \int dR' \rho_F(R_1,R';t-t')K(R',R_2;t') \quad (4.4)$$

where the kernel is non-zero only on the nodes of ρ_T where it equals the discontinuous derivative of $\tilde{\rho}_T$ across the node.

$$K(R_1,R_2,t) = \lambda\delta(\rho_T(R_1,R_2;t)) \nabla[\tilde{\rho}_T(R_1^+,R_2;t) - \tilde{\rho}_T(R_1^-,R_2;t)] \quad (4.5)$$

Before we discuss forms for the trial density matrix, let us return to the more general case where one has spin $\frac{1}{2}$ fermions. Assume that there is no magnetic field so that the Hamiltonian is independent of spin. Then we can quantize the spin axis in the $\hat{z}$ direction. Suppose we want to calculate the partition function Z_m in an ensemble where S_z is fixed to be m. It is not hard to see that this equals:

$$Z_m = \frac{1}{N!} Tr_\sigma \left\{ \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \delta^N(\sigma,\mathcal{P}\sigma)\delta(\sum_i \sigma_i, m) \int dR \rho_D(R,\mathcal{P}R) \right\} \quad (4.6)$$

Now simply relabel the coordinate indices in the integration so that particles 1 through $N_\uparrow = N/2 + m$ have up spin and particles $N_\uparrow + 1$ through N have down spin. Then clearly all arrangements of the spin variable in the trace contribution the same amount to the partition function and we can just calculate one term and multiply by the number of ways of assigning spins. There are $N!/(N_\uparrow! N_\downarrow!)$ such arrangements. Define the following operator to antisymmetrize over the up spins and the down spins individually.

$$\mathcal{A} = \frac{1}{N\uparrow! N\downarrow!} \sum_{\mathcal{P}_\uparrow,\mathcal{P}_\downarrow} (-1)^{\mathcal{P}_\uparrow+\mathcal{P}_\downarrow} \mathcal{P}_\uparrow \otimes \mathcal{P}_\downarrow \quad (4.7)$$

where $\mathcal{P}\uparrow$ ($\downarrow$) operates on the up (down) spin coordinates only. With this one can calculate spin-independent correlation functions in the fixed S_z ensemble.

If we want to calculate the total partition function: $Z = \sum_m Z_m$ the magnetization must be sampled. To average over all spins one could occasionally attempt a spin flip from one fixed m ensemble to $m \pm 1$ and accept or reject that flip, depending on whether the path is now legal.

We can also do a fixed S ensemble by using a different projection operator. For example one can use the relationship:

$$Z_S = (2S+1)(Z_m - Z_{m+1}) \quad (4.8)$$

and try to calculate the differences of partition functions. Equivalently one could use a projection operator for the total spin. This is discussed in detail in the recent work of Lyubartsev and Vorontsov-Velyaminov (1993). As an example, suppose we want to calculate the partition function restricted to the spin-singlet states of a four electron system like the Be atom. Then one wants to sum only over 12 of the total number of 4!=24 permutations. The identity permutation gets a relative weight of 2, the three possible double pair exchanges such as (2,1,4,3) get a relative weight of 2, and the eight possible triplet exchanges such as (2,3,1,4) get a relative weight of -1. These weights specify the projection operator that is used for the initial value of the density matrix. What is not clear to me is how these approaches will scale in the number of fermions.

V. THE REFERENCE POINT

We call R_* the reference point and it plays a very special role in restricted path integrals since it is the value of the density matrix with respect to the reference point that restricts the paths. Averages such as the density can only be taken at the reference point. This is different from boson or distinguishable path integrals; for them all time slices are equal by an obvious time-slice symmetry, and all time slices are equivalent in taking averages. If we write expectation values in terms of derivatives of the partition function (e.g. the kinetic energy is the mass derivative of the partition function), one recovers an expectation value involving all the time slices, not just one containing the reference point. However, evaluating the derivatives can be quite complicated.

By a "time-independent" or "ground-state" restriction is meant that the restriction does not depend on the reference point. This is achieved by using an antisymmetric trial wavefunction $\Psi_T(R)$ and requiring that $\Psi_T(R_t) \neq 0$ throughout the path. This is identical to the ground state fixed-node method (diffusion or Green's Function Monte Carlo). The algorithm is considerably simpler than using time-dependent nodes and time-slice symmetry is restored. However, calculations with ground-state nodes on liquid ^{3}He gave a poorer description of the properties of the liquid at non-zero temperature (Ceperley, 1992). It is not clear in what sense this time-independent restriction would give proper finite-temperature effects, even though one is using path integrals. I speculate that any exact fermion description at finite temperature must break the time-slice symmetry if it is to solve the sign problem.

There is a simple trick for using two reference points instead of one. This improves the method in several ways. Let us square Eq. (4.2). This means we construct two paths, one from R_0 to R_β and another from

R_0' to R_β both in time β. The weight of the product of the two paths will be: $\rho_F(R_0, R_*; 0)\rho_F(R_0', R_*; 0)$. When we have integrated over all paths, and the two initial positions R_0, R_0' and the reference point, we will have determined the density matrix $\int dR_* \tilde{\rho}_T(R_*, R_\beta; \beta)^2 \approx \rho_F(R_\beta, R_\beta; 2\beta)$, that is at a temperature half as big. Note that both of the paths are node avoiding with respect to the same reference point, R_* and end up at the same point R_β at the same time. Hence, the sign of $\rho_F(R_0, R_*; 0)$ equals the sign of $\rho_F(R_*, R_\beta, \beta)$, which equals the sign of $\rho_F(R_0', R_*; 0)$. The contribution of the double path is positive. Also note a point $\mathcal{P}R_*$ has the same effect on the restriction as does R_* since $\rho_F(R, R_*) = 0$ implies that $\rho_F(R, \mathcal{P}R_*) = 0$. This means we can work with only a single projection and use a weight for the double path since:

$$\rho_F(R_0, R_0'; 0) = \int dR_* \rho_F(R_0, R_*; 0)\rho_F(R_0', R_*; 0). \quad (5.1)$$

The net effect of doubling the time is to redefine how we measure time differences in the restriction. Let the reference point be defined as having "time" value of zero. Then the time argument that we should use for the restriction (the restricted time) is:

$$t_r = \begin{cases} t & \text{for } 0 \le t \le \beta/2 \\ \beta - t & \text{for } \beta/2 \le t \le \beta \end{cases} \quad (5.2)$$

(Now we are using an inverse temperature β ***not*** 2β.)

Time doubling is an improvement because if we have accurate nodes down to a temperature T, we can do accurate simulations down to $T/2$. Also, we can use both R_0 and $R_{\beta/2}$ to calculate expectation values. Finally, we have restored (imaginary) "time-reversal" invariance. If our nodes are incorrect, then in general: $\tilde{\rho}_T(R_1, R_0; \beta) \neq \tilde{\rho}_T(R_0, R_1; \beta)$; the symmetry of Eq. (2.7) could be violated. By using two paths from R_0 to $R_{\beta/2}$ we are performing the integral:

$$\tilde{\rho}_T(R_1, R_2; \beta) = \int dR_* \tilde{\rho}_T(R_1, R_*; \beta/2)\tilde{\rho}_T(R_2, R_*; \beta/2) \quad (5.3)$$

which is clearly symmetric in R_1 and R_2.

The time doubling cannot be repeated without reintroducing the sign problem. Also, one gets a sign problem if the two time arguments that are multiplied together are different, so this method cannot be used to determine imaginary time correlation functions without having sign problems. But it is well worth doing it once.

VI. AN EXAMPLE OF RESTRICTED PATHS

To get a feeling for restricted paths let us consider the problem of molecular hydrogen. We will work in the Born-Oppenheimer approximation so the two protons in a single hydrogen molecule are represented by two spin 1/2 particles interacting with an attractive potential. The total spin (S) and total orbital angular momentum (L) are good quantum numbers. The spin 0 state is called para-hydrogen, and must have an even value of L to keep the molecular wavefunction antisymmetric in spin and coordinates. The spin 1 states are called ortho-hydrogen and they must have odd values of L. Often the hydrogen cannot easily change its spin state, so that para- and ortho-hydrogen can be considered as separate chemical species, for a time at least. They can change their angular momentum values with collisions with other molecules but not easily their spin. A third possibility of statistics is if the two nuclei are different particles, ***e. g.*** a proton and a deuteron, in which case they obey distinguishable particle or Boltzmann statistics.

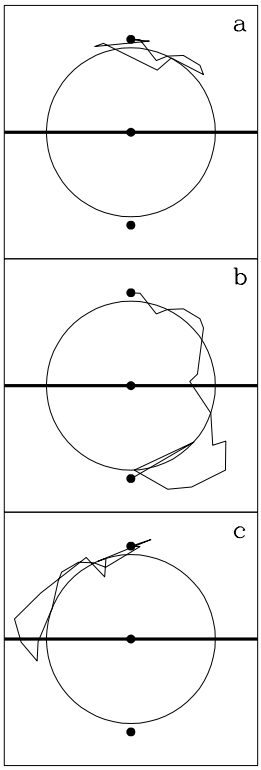

FIG. 2. Three types of paths that contribute to the partition function of a single hydrogen molecule. Shown is the vector separation between the two atoms; the sphere represents the repulsion when the two atoms approach each other. The two small circles, represent the reference point, the origin of time. The irregular curve represents the path; it starts and ends at a reference point. The horizontal line represents the nodal plane in the ortho-hydrogen state.

Let us suppose that the particles are massive enough that the relative coordinate is almost fixed at a given radius $r_0 \approx 0.75$ Å. Hence the relative coordinate $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ is almost fixed on the surface of a sphere. In figure (2) we show four types of paths. The reference point is the large dot.

Now consider what we need to do to calculate the partition function for the three types of statistics. Distinguishable particles are the simplest: allow all paths returning to the starting point (type A in the figure). For ortho- and para-hydrogen, we can use parity to project out the correct states. This generates paths of type B which end up at the opposite pole in relative coordinates. For para-hydrogen the direct method would be to sum over all paths of types A and B. Ortho-hydrogen would be to sum over paths of type A but subtract the contribution from paths of type B.

In the case of ortho-hydrogen it is easy to calculate the exact nodes of the density matrix. In relative coordinates any wavefunction has the angular factor $Y_{lm}(\hat{\mathbf{r}})$. Then in the sum over the quantum states for different m in Eq. (2.2), using the addition formula for spherical harmonics we obtain a factor $P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_*)$. Since all the odd Legendre polynomials vanish when their arguments vanish the ortho-density matrix vanishes when $\mathbf{r} \cdot \mathbf{r}_* = 0$. In general, additional nodes would be possible, but the hydrogen molecule has only this one planar node. Paths of type C are node-crossing as opposed to the node-avoiding path A.

To summarize, we add the following classes of path for the different statistics of the hydrogen molecule:

1. Distinguishable hydrogen: A+C
2. Para-hydrogen: A+B+C
3. Ortho-hydrogen (direct method): A-B+C
4. Ortho-hydrogen (restricted method): A only ($\mathbf{r}_t \cdot \mathbf{r}_* > 0$)

The reason that restricted path integrals give the same value is that paths of type B and C can be paired together and canceled off against each other. This is because the flux of paths is the gradient of the density matrix at the node and since the gradient is continuous across the node, the positive paths crossing at a given nodal point will precisely cancel against the negative paths.

Hence the restricted paths are limited to be in a half-space. Note that there is no definite location of this half-space. Its position depends on the reference point, even at zero temperature. This is because the ground state of S=1 is three-fold degenerate. Isotropy is restored by averaging over the reference point position.

Consider now a crystal made of hydrogen molecules. To a good approximation one can neglect the angular coupling of the spins in constructing the nodal surface. From the point of view of restricted path integrals, ortho-hydrogen is more easily orientable than para-hydrogen because its restricted paths are naturally dumb-bell shaped. Runge et al (1992) have used this approach in PIMC simulations of both ortho- and para-hydrogen. On the other hand, Buch (1994) in his PIMC calculations, explicitly diagonalized the rotational Hamiltonian at each step, a considerably more complicated and more approximate way of handling ortho-hydrogen.

VII. NODES OF THE DENSITY MATRIX

The only uncontrolled approximation in the Restricted Path Integral Method is the restriction, the rule by which we allow paths. Clearly the success of the method hinges on the choice of this restriction. The situation is not very different from that of classical Monte Carlo or Molecular Dynamics simulations. Except for a few isolated cases, the true classical Hamiltonian describing a real physical system is not very well known (say compared to a typical thermal energy of 100 K $\approx$ 0.01 eV $\approx$ 0.35 mH). I am not referring to simplified models like the Ising model or a Lennard-Jones liquid but about water or a polymer. In principle, the classical dynamics is controlled by the true Born-Oppenheimer potential. In practice, a very simplified potential is used, either a semi-empirical one or one from on-the-fly LDA calculation as Car has described. One knows that the true trajectory deviates very quickly from the calculated trajectory and that detailed properties such as the melting point depend on the actual potential. However, simulations are still useful because they are true simulations of a nearby model system. Hopefully, the model is in the right universality class. Simulations of good models are useful because they are the only way to understand reliably many-body effects. Even if some of the quantitative details are inaccurate, if we can characterize the nodal restriction sufficiently well, the simulations will be useful in understanding nearby strongly-interacting fermion systems.

Let me explore further this analogy between classical simulations and RPIMC. The dimensionality of the unknown "potentials" is not so different. The Born-Oppenheimer classical potential is a real function of $3N$ variables. The nodal restriction is a Boolean function (i.e. only the sign matters) of $6N + 1$ variables (counting the two legs of the density matrix and imaginary time). Another similarity is that at finite temperature, the nodal surface is local. By this we mean that the nodal surface is determined by the equality of even and odd local permutations; local in that each particle can only move on the order of a thermal wavelength. For classical neutral atoms with localized electrons, one generally assumes the potential is a pairwise sum with smaller and higher three-body contributions, so the many-body potential can be

built up from the properties of small clusters. In both cases, permutational and other symmetries can be used to reduce drastically the phase space where we need to determine the "potentials."

The nodal surfaces have hardly been investigated, so there is much to be done here. It is my opinion that we should not be put off by the RPIMC method just because we do not know the exact restriction, but we should take inspiration from what has happened in classical simulations and push ahead with a nodal ansatz, all the while trying to improve our knowledge of fermion nodes. Hopefully, the trial nodes are good enough to put the model in the right universality class. What we don't have is the century of accumulated wisdom about what nodal properties are important. In the rest of this section, we review what is known about the nodes, and various approximations that could be used.

In an actual calculation, one does not make a geometric interpretation of the nodes as we did in the two-particle example. Instead one computes a ***trial density matrix*** and uses that to decide whether a given path is to be allowed. Paths for which the trial density matrix are negative or too close to zero are rejected. I will discuss this aspect in the next section. How can we choose a fully antisymmetric trial density matrix which can be quickly computed during the Monte Carlo random walk? Using the projection, the fermion density matrix is:

$$\rho_F(R,R_*;t) = \mathcal{A}\rho_D(R,R_*;t) \propto \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \rho_D(\mathcal{P}R,R_*;t). \tag{7.1}$$

Only the sign matters so we can drop any constants. This formula is exact but not useful unless a) we have a good expression for ρ_D down to temperatures as low as $2T$ and b) we can evaluate quickly the permutational sum. Effectively, the only convenient antisymmetric density matrices for the many-body case are determinants of one form or the other.

For the moment, let us consider the reference point R_* and the inverse temperature t as fixed parameters. Then the nodes have dimension $3N-1$ since a single equation specifies whether R_t is on a node. One property that holds true in general, is that when two fermions have both the same spin and the same spatial coordinates, all the wave functions and hence the density matrix must vanish. Hence, for any pair of fermions with the same spin, the hyperplane defined by the equation: $\mathbf{r}_i = \mathbf{r}_j$ is on the node. Since these are three equations (in three dimensions) the "coincident hyper-planes" have dimensionality $3N-3$. The coincident planes are fixed hyper-points lying on the nodal surfaces which have a dimensionality two larger. For quantum mechanics in one dimension, the coincident points exhaust the nodal surfaces so that one knows the exact restriction. This has misled many researchers into believing that the situation in higher dimension was not much more difficult. Such is not the case, at least as far as we know. For fermions in two or three dimensions, symmetry is not sufficient to determine the position of the nodes. Their position depends in a non-trivial way on the potential. (There is also a significant difference in complexity between 1 and higher dimensions. In 1D the sign of the density matrix is equivalent to ordering on a line, so computing the restriction is a completely local evaluation, while in higher dimension computing even the free particle density matrix costs $\mathcal{O}[N^3]$ operations.) Knowing the coincident points are on the nodes means that if one particle is displaced roughly an interparticle spacing, typically one will cross a node. Later, we will see the connection between the nodes and the momentum distribution.

Now, let us discuss the nodal surfaces of non-interacting fermions. Let $v_e(\mathbf{r})$ be a single-particle external potential. The distinguishable particle density matrix is then a product of solutions of the single-particle Bloch equation:

$$-\frac{dg(\mathbf{r},t)}{dt} = [-\lambda\nabla^2 + v_e(\mathbf{r})]g(\mathbf{r},t) \tag{7.2}$$

with the boundary condition:

$$g(\mathbf{r},\mathbf{r}_*;0) = \delta(\mathbf{r}-\mathbf{r}_*). \tag{7.3}$$

Then using the antisymmetric projection operator and the definition of a determinant, we find for the spinless case:

$$\rho_F(R,R_*;t) = \frac{1}{N!} det[g(\mathbf{r}_i,\mathbf{r}_{j,*};t]. \tag{7.4}$$

In the case where the external potential is zero (or a constant), the single-particle density matrix is a Gaussian.

$$g(\mathbf{r},\mathbf{r}_*;t) \propto \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_*)^2}{4\lambda t}\right]. \tag{7.5}$$

Note that in periodic boundary conditions, the relative distances should be computed using the minimum image convention. But if the thermal wavelength $\sqrt{2\lambda t}$ approaches the size of the box, which can occur at low temperatures, additional images may be required. The exact single-particle density matrix is a theta function in periodic boundary conditions. Only if the exact density matrix is used in the determinant, will the many-body density matrix approach the Slater determinant of plane waves appropriate to the ground state of a uniform liquid. (This is an issue for restricted fermion path integrals but not for boson path integrals because the "time" argument ranges up to $1/(2T)$ instead of just up to $\tau = 1/(MT)$.)

At high temperature these nodes have a geometrical construction. Consider sitting at a point R and taking

the limit as $t \to 0^+$. For a typical point, a single permutation $\mathcal{P}R_*$ will be closest to R and all the other permutations can be neglected. If $\mathcal{P}$ is even, ρ_F will be positive, otherwise it will be negative. A point R is on a node if it is precisely equidistant from an even permutation and an odd permutation.

$$(R - \mathcal{P}_e R_*)^2 = (R - \mathcal{P}_o R_*)^2 \tag{7.6}$$

or

$$R \cdot (\mathcal{P}_e - \mathcal{P}_o) R_* = 0. \tag{7.7}$$

Hence at high temperatures the nodal surfaces are hyperplanes. A nodal region is the set of points closer to R_* than to any other odd permutations $\mathcal{P}R_*$. In this sense, the nodes are as far away from the reference point as an antisymmetric function will allow.

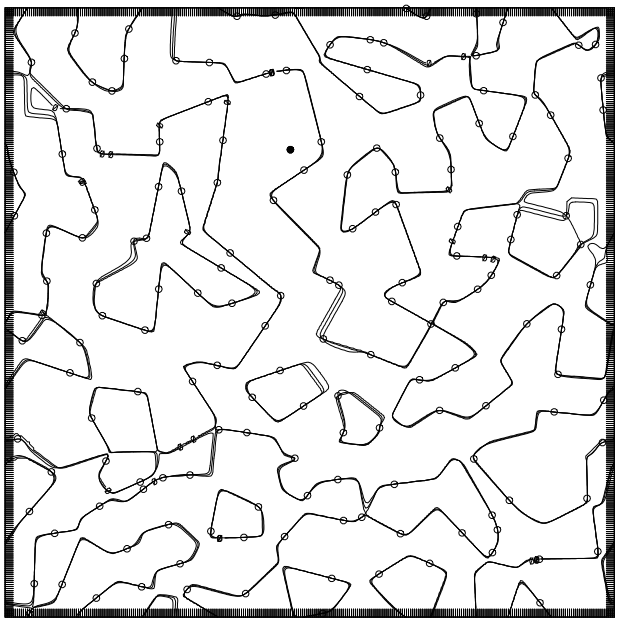

FIG. 3. A 2D cross section of the density matrix of 161 free spinless fermions in a periodic square. The open circles are the fixed positions of 160 fermions;the last fermion is free to roam around the squares. The contour lines are the nodes of the density matrix. The curves respresent the nodes for temperatures in the range 2.5 $E_F \le T \le 10E_F$ where E_F is the Fermi energy.

On the other hand, at low temperatures the free-particle nodes are determined by minimizing the kinetic energy. They must be as smooth as possible while keeping the coincident points on them. Hence the nodes change character as the evolve from high temperature to low temperature as seen in Fig. (3-4).

I have investigated (Ceperley, 1991) the number of different nodal regions into which the full configuration space is divided. I have found that the free-particle spinless-fermion density matrix in periodic boundary conditions divides the total phase space into only a single positive region and a single negative region (except in one spatial dimension where the nodes divide the phase space into $N!$ regions.) This means that the restricted path partition function includes contributions from all $N!/2$ even permutations.

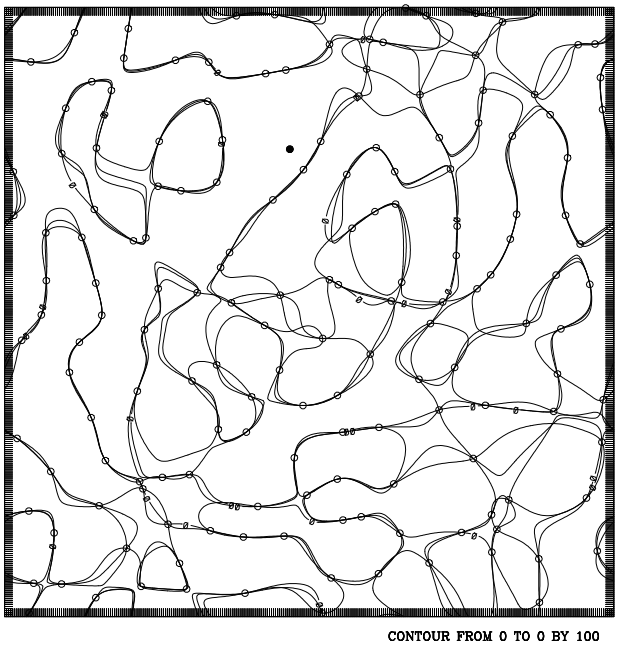

FIG. 4. A 2D cross section of the density matrix as in the previous figure. In this figure the temperature is close to the ground state; the temperature ranges from 0.04 $E_F \le T \le 0.25E_F$.

To go beyond free particle nodes, we can first work within a one-particle self-consistent approach, such as Hartree-Fock. This has the virtues of a) giving a Slater determinant of single particle density matrices which will be easy to evaluate and b) giving good nodal surfaces at zero temperature. It has been found that nodes coming from Hartree-Fock calculations and LDA calculations work quite well for electronic systems. All that is required in the above equation (7.2) is to use an effective external potential instead of the bare potential. It is possible to have the effective potential be refer-

ence point dependent $v(\mathbf{r}, \mathbf{r}_*; t)$ and for the mass be an effective mass which is "time-dependent" $\lambda(t)$. This roughly corresponds to unrestricted Hartree-Fock, or to self-interaction corrected density functional theory.

Another approach is to take what we know about good distinguishable particle density matrices and directly apply formula Eq. (7.1). The pair product density matrix is:

$$\rho_D(R, R_*; t) = \rho_{D0}(R, R_*; t) exp[-\sum_{i<j} u(r_{ij}, r_{ij,*}; t)] \tag{7.8}$$

where the pair action u is defined either in terms of the exact action of a pair of particles or as the first cumulant action. This form is quite accurate for systems interacting with pair potentials (Ceperley, 1995) at least above the degeneracy temperature.

In the simplest approximation we make the end-point approximation for the pair action: $u(r, r_*; t) = \frac{1}{2}(u_d(r; t) + u_d(r_*; t)) + \mathcal{O}[t^2]$. But since the total pair action is symmetric under permutations, the pair terms will factor out of the antisymmetrization and hence will not affect the nodes of the free particle density matrix. This explains why the nodes of the free-particle density matrix, even though it is over-simplified are a reasonably good approximation down to the degeneracy temperature. The off-diagonal corrections, which could push the nodes away from the free-particle positions, are $\mathcal{O}[t^2]$, while the dominant kinetic energy term is $\mathcal{O}[t^{-1}]$. The temperature has to be fairly low for contributions coming from the potential to matter. This is supported by a perturbation calculation of Hall and Price (1991).

An important feature of accurate actions is that they contain off-diagonal pair terms, *i.e.* the pair action depends on both R and R_* in a more complicated way than this. But without the end-point approximation, the resulting antisymmetric function is not a determinant, so that can only be computed directly for systems with a few fermions where one can explicitly sum up the permutations. One pathological feature of free-particle nodes is that there is no coupling in the nodes with unlike spins. The nodal surfaces are a Cartesian product of up-spin nodes with down-spin nodes. This feature is incorrect and possibly very significant in a strongly-coupled system such as liquid ^{3}He where collective spin excitations are important. [A good homework problem is to verify that in a four electron system such as the Be atom, the restricted paths allow not only the identity permutation, the only one allowed in Hartree-Fock, but also the double pair permutation (2,1,4,3).]

In general, we will have to include the effects of the off-diagonal terms into the diagonal ones. The principle physical effect of the potential is to slow the free-particle diffusion. One can define an effective (inverse) mass by:

$$\lambda^*(t) = \langle (\mathbf{r}(t) - \mathbf{r}_*)^2 \rangle / (6t) \tag{7.9}$$

where the average is over distinguishable particle paths starting from the point $\mathbf{r}_*$. This effective mass is computed in Ceperley (1995) for liquid ^{4}He as a function of "time". At short time one has the bare mass but at larger times (when the thermal wavelength equals the interparticle spacing) the effect of the interparticle interaction is to slow down the diffusion by a factor of two. This will not affect the nodes at zero temperature, but it will rescale the temperature at which those nodes are used. In a localized phase such as a solid helium, the effective mass diverges so that the single particle density matrix remains localized even at zero temperature.

Let us take a least-squares function-fitting approach and say the best single particle function to represent the many-body density matrix is the one which minimizes the mean squared action:

$$\chi^2 = \int dR dR_* \mu(R, R_*; t) \left[S_D(R, R_*; t) - S_A(R, R_*; t) \right]^2. \tag{7.10}$$

Here $\mu(R, R_*; t)$ is an measure for the end-points of a path chosen to generate a good "sampling" of configuration space. For example, one can take pairs of points (R, R_*) from a RPIMC simulation of the system in questions and relabel the R from the distribution of permutations $\rho_D(\mathcal{P}R, R_*; t)$. In the above equation, $S_D(R, R_*; t)$ is the exact distinguishable action

$$S_A(R, R_*; t) = \frac{1}{2}[U_d(R; t) + U_d(R_*; t)] + \sum_{i=1}^{N} s(\mathbf{r}_i - \mathbf{r}_{i*}; t) \tag{7.11}$$

is the general form of an approximate action which maps into a determinant after antisymmetrization and $U_d(R; t)$ is the exact diagonal action. Minimizing the rms action with respect to the function $s(\mathbf{r}, \mathbf{r}_*; t)$, we find the optimal single particle density matrix is proportional to :

$$s(\mathbf{r}_1, \mathbf{r}_{1*}) = \langle S_D(R, R_*) - \frac{U_d(R) + U_d(R_*)}{2} - \sum_{i=2}^{N} s(\mathbf{r}_i - \mathbf{r}_{i*}) \rangle_{\mathbf{r}_1, \mathbf{r}_{1*}} \tag{7.12}$$

where $\langle \ldots \rangle_{\mathbf{r}_1, \mathbf{r}_{1*}}$ means an average over μ with $(\mathbf{r}_1, \mathbf{r}_{1*})$ held fixed. This equation can be solved by iteration: assume the free-particle value for $s(\mathbf{r}, \mathbf{r}_*; t) = (\mathbf{r} - \mathbf{r}_*)^2/(4\lambda t)$ evaluate the left-hand-side, and iterate. This will reduce to the effective mass formula above if we assume the single particle density matrix is Gaussian and we assume the action is normally distributed.

A more complex way of changing the nodes is to put in backflow effects. This has been found to be very successful for liquid ^{3}He (Panoff and Carlson, 1989) and in

the 2D electron gas (Kwon, Martin and Ceperley, 1993) getting on the order of 90% of the energy missed by the free-particle nodes. To compute the Slater determinant, one first transforms the coordinates of all the particles on the path to "quasi-particle" coordinates. The simplest transformation allowed by translation symmetry is:

$$\mathbf{s}_i = \mathbf{r}_i + \sum_{j \neq i} (\mathbf{r}_i - \mathbf{r}_j)\eta(|\mathbf{r}_i - \mathbf{r}_j|; t) \tag{7.13}$$

where $\eta(r;t)$ is the "backflow" function. Then the trial density matrix is the usual Slater determinant but written in quasi-coordinates:

$$\rho_T(R, R_*; t) = det[g(\mathbf{s}_i, \mathbf{s}_{j*}; t)]. \tag{7.14}$$

At high temperature one can show from the Feynman-Kacs formula that the backflow function will depend on the potential as:

$$\lim_{t \to 0} \eta(r,t) = \frac{\lambda t^2 dv(r)}{2r dr} \tag{7.15}$$

Hence at high-temperature, the classical force pushes the nodal surface by an amount proportional to t^2. At low temperature the backflow function $\eta(r)$ saturates to its zero-temperature form. The difficulty with using a backflow trial density matrix is that all the coordinates get coupled together, so that updates of the determinant are much more expensive. This is the reason backflow nodes have not been often used, even in the ground state. In addition, there has been little investigation of ways of obtaining the backflow function.

As we mentioned above in the two particle example, for the exact nodal surface the positive and negative paths push the nodes equally hard from the two sides. There is a sort of equality of nodal surface pressure across the nodal membrane. Unfortunately this does not lead to a practical general way of constructing nodes because the membrane is defined in such high dimensionality. It would take too many paths to do that.

We must bear in mind that all this talk about nodal approximation should not make us lose sight of the fact that correlation caused by the potential is being exactly treated with the distinguishable path integrals. It is only the shape of the playing field, not its size or the playing rules, that we that we must approximate. Ground state calculations of such electronic systems as small molecules, the electron gas or many-body hydrogen, show that the fixed-node calculation using Hartree-Fock nodes have an absolute accuracy of better than 0.01 eV/molecule. Relative accuracies are expected to be significantly better. One might also suspect that the restricted path errors would increase with decreasing temperature, so that the ground state would be be worst case.

VIII. SOME TECHNICAL DETAILS OF RPIMC

A. The action

The action for restricted fermions, which we write as $S_F(R_t, R_{t-\tau}; \tau; t_r, R_*)$, is more complicated and depends on more variables than the distinguishable action S_D because of the lack of time symmetry in the paths. It is a function of R_* and t in addition to its usual dependence on $R_t, R_{t+\tau}$ and τ. Usually we will not write these extra variables, taking them to be understood. For ground state nodes (no time dependence) one reverts to the bosonic situation; the action does not depend on the reference point or reference time. The Feynman-Kacs formula, Eq. (2.15), for the action arising from the potential and the restriction is:

$$e^{-U_F(R_t, R_{t+\tau})} = \left\langle exp\left[-\int_t^{t+\tau} dt' V(R_{t'})\right]\right\rangle_{R_t \to R_{t+\tau} \in \Upsilon(R_*)} . \tag{8.1}$$

The averaging is over all Brownian walks starting at R_t and ending at $R_{t+\tau}$ and keeping inside $\Upsilon(R_*)$.

The ***primitive-nodal*** action makes the approximation of checking the restriction only at the M sampled points on the path. We only check to see whether $\rho_T(R_j, R_*; t_j) < 0$ for $0 \leq j \leq M$. To illustrate the problem with the primitive nodal action, consider a single particle in a 1D well of size a:

$$V(x) = \begin{cases} 0 & \text{for } 0 < x < a \\ +\infty & \text{otherwise} \end{cases} \tag{8.2}$$

In this model, the nodes are fixed, as they would be for any fermion problem at low temperatures. The error of the ground state energy, ($\beta = \infty$) is shown in figure 5 versus the time step. In the primitive approximation the error decreases as $\tau^{1/2}$. This dependence is easy to understand. The energy of a box of size a is $\lambda(\pi/a)^2$. Using the primitive approximation effectively increases the size of the box by an amount proportional to the thermal de-Broglie wavelength of a time step: $(\lambda\tau)^{1/2}$. Shown in figure 6 is the density computed with the primitive action. The primitive-nodal action allows the density to be non-zero at the nodes. Hence, the energy is decreased by a relative amount $(\lambda\tau)^{1/2}/a$. In boson problems, we can find actions for which the error in the energy converges much faster, as τ^3, so that use of the primitive approximation will seriously degrade the overall performance of the PIMC algorithm.

Luckily, one can do considerably better, so that fewer time slices are needed to accurately represent the path. The action picks up a contribution from the nodes because walks can wander back and forth across the nodes even though they happen to be on the correct side at the

sampled points, R_t and $R_{t+\tau}$. Improved fixed-node sampling methods have been developed for the ground state simulations by Anderson (1976). Here we adapt this to path integrals. To systematically improve the action, we have to make a sequence of three approximations.

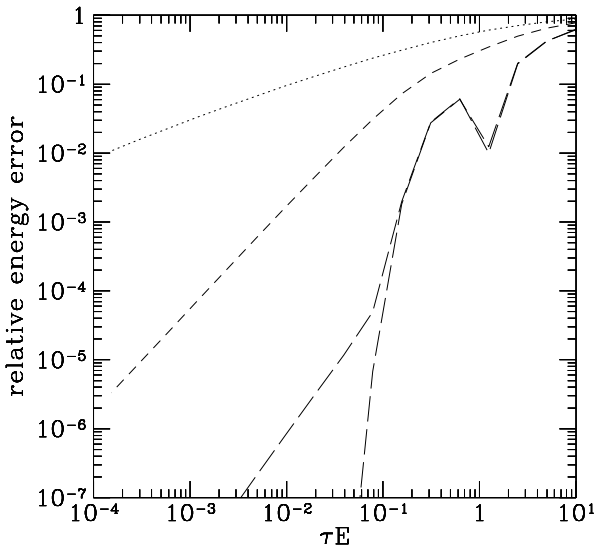

FIG. 5. The energy error of a particle in a square well with a harmonic potential of size π at zero temperature computed with four different forms for the action as a function of time step. The actions differ in their estimate of the distance d of the particle to the wall. The top dotted curve is for the primitive action, the distance is set to infinity, the error decreases as $\tau^{0.5}$; the next lower short dashed curve uses the newton estimate with the nodal distance estimated from the exact ground state wavefunction so that $d = \tan(x)$, its error decreases as $\tau^{1.5}$. In the next lower dashed curve, the exact nodal distance is used but no account is made for the coupling between the bosonic action and the nodal action. In the lowest dashed curve the distance is estimated as $d = min(x, \pi - x)$, and there is no bosonic action, the error is due to neglect of the mutiple images goes as $exp(-\pi^2/4\tau)$

In first approximation to the average of Eq. (8.1), we assume that the potential energy is uncorrelated with crossing and re-crossing of the nodal surface. Then the action factorizes

$$\exp[-U_F(R_t, R_{t+\tau})] = \exp[-U(R_t, R_{t+\tau}) - U_N(R_t, R_{t+\tau})] \tag{8.3}$$

where U is the purely boltzmannon action and U_N is the free particle action due to just the restriction. The second to lowest curve in fig. (5) shows the effect of this approximation on the ground state energy for a particle in a well with an additional harmonic potential. The error in the energy caused by the factorization goes as τ^2. The boltzmannon action U has been extensively investigated for use in the path integrals (Ceperley, 1995) so we need say no more about it here.

The exact nodal action for a particle in a box is easy to calculate. In the case of a particle confined only to the half-space $x > 0$, one can solve the Bloch equation by the method of images. The total density matrix is the difference: $\rho(x, x') - \rho(x, -x')$ since the difference satisfies the Bloch equation and the boundary condition at $t = 0$ and $x = 0$. Using the form for the free particle action the nodal action is:

$$U_N(x_t, x_{t+\tau}) = -\ln[1 - exp(-\frac{d_t d_{t+\tau}}{\lambda\tau})] \tag{8.4}$$

where d_t is the distance of x_t to the node at time t and $d_{t+\tau}$ is the distance of $x_{t+\tau}$ from the node at time $t + \tau$. Hence the action diverges logarithmically near the node and is significantly repulsive in a region on the order of $\sqrt{\lambda\tau}$. See figure 7.

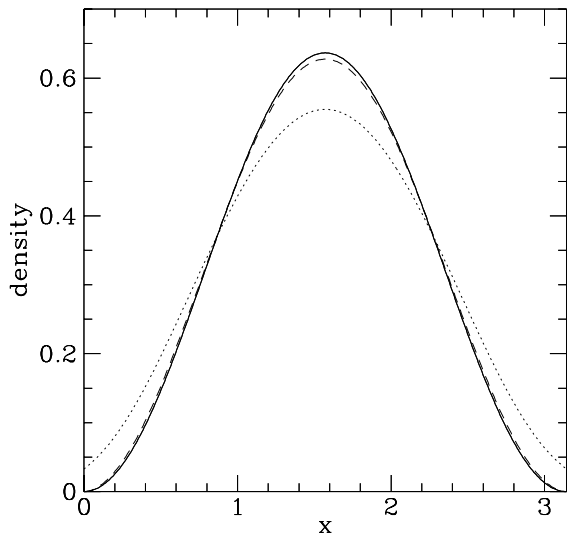

FIG. 6. The single particle density in a square well at zero temperature with size π computed with two forms for the action for a time step of $\tau E = 0.078$ (the corresponding de Broglie wave length 0.4). The solid line is the exact density $\sin(x)^2$. The dotted line is the primitive approximation, its density is not zero at the edge of the box. The dashed line was computed using the image action. The corresponding energies are shown if Fig. 5.

In particle-in-the-box model, there are boundaries (nodes) both at $x = 0$ and at $x = a$. Hence, the above action is not exactly correct. As in similar problems in electrostatics, one must expand in an infinite series of

positive and negative images. In the case of the density matrix, the images decay as $\exp(-L^2/(4\lambda\tau))$, so that at small τ the random walk sees only the nearest wall. If we make the approximation of only using the nearest image, we get the error shown as the lowest line in figure 5. Its error converges exponentially in τ. The dominant error at small time step is the neglect of the potential, not the neglect of the other walls.

Now, let us return to the many-body case. The free-particle nodal action is difficult to evaluate (at least I do not know how to do it), so we need to make further approximations. We shall assume that locally the nodal surface is constant (as a function of t) and is a single hyper-plane so we can use the image approximation discussed above.

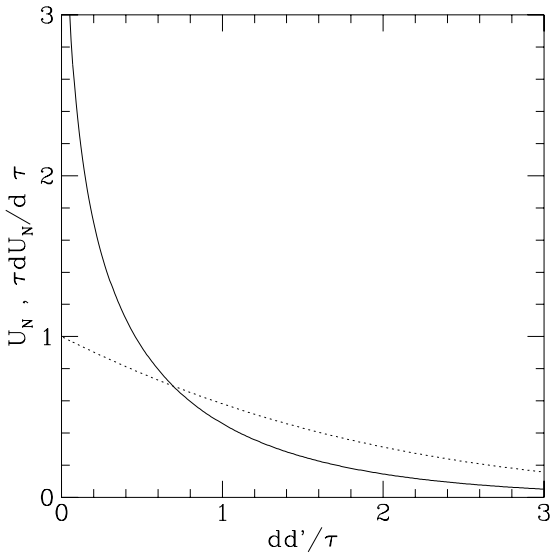

FIG. 7. The action (solid line) and the energy ($\tau dU/d\tau$), dashed line, for a 1D particle next to a wall at the origin.

The remaining non-trivial problem is how to estimate the distance d to the node. In the many-body time-dependent case we define the distance to the node as $d_t = \min(|R_n - R_t|)$ where R_n varies over all points with $\rho_F(R_n, R_*; t) = 0$. It is easy to determine an upper bound to d_t. If any two particles are moved directly towards each other, they will be on a node when they coincide. It follows that $d_t \le r_{ij,t}/\sqrt{2}$. However the distance may be much smaller than this. To get an estimate, we can use the Newton-Raphson method: given a function (hopefully smooth) $f(R)$ which vanishes when $\rho_T(R)$ does, an estimate of the nodal distance is:

$$d(R) \approx \frac{|f(R)|}{|\nabla f(R)|}. \tag{8.5}$$

This is good as long as the contribution of higher order derivatives of f is not large within a thermal wavelength.

For the particle-in-a-box model, we took $f(x) = \sin(\pi x/a)$ for the purposes of ascertaining the effect of estimating inaccurately the nodal distance. This gives an estimate of the nodal distance $d = (a/\pi)|\tan(\pi x/a)|$. The energy with this assumption is shown as the middle dashed line: the error converges as $\tau^{1.5}$. The action is worse because the precise distance of the walk from the node is not known. Nonetheless, the error is much superior to that using the primitive action.

In practice, one should choose f differently in the low temperature and high temperature limits, as shown in figure (8). At low temperatures, as we have discussed, the nodes are smooth and relatively independent of R_* and t. Then we choose $f(R) = \rho_T(R, R_*; t)$. On the other hand for small values of t the density matrix is very sharply peaked. The curvature of Gaussian density matrix always places the distance too close, so we consider an alternative procedure where one divides out the nearest Gaussian. The idea is to make f as linear as possible. Define $f(R) = \rho_T(R, R_o; t)/\rho(R, \mathcal{P}R_*; t)$ where $\mathcal{P}$ is the permutation having the largest term in the determinant: the permutation which minimizes $\sum_i (r_{i*} - r_{\mathcal{P}_i})^2$. To determine the best permutation, we use the Hungarian algorithm for solving the linear sum assignment problem (Burkard and Derigs, 1980). That takes $\mathcal{O}[N^3]$ operations.

Although we have made lots of approximations in deriving the nodal action, it is correct in the limit that $d_t \to 0$. This is an essential property of any approximation. It matters little what you do for particles far away from the nodal surface. At high temperature, the nodes are far away (it some sense that are as far away as they can be) so the upper bound is reasonable. At low temperature, the wave function and nodes are smooth and planar, see figure 4. There the Newton-Raphson procedure works well.

Now consider the problem of how to calculate the internal energy. The simplest way of calculating the energy is to take the time derivative of the action. This is called the thermodynamic estimator in Ceperley (1995). The direct fermion contribution to the energy of the nodal action is:

$$E_N = \left\langle \frac{dU_N}{d\tau} \right\rangle = \sum_{i=1}^{M} \left\langle \frac{x_i}{\tau(e^{x_i}-1)} \left\{ 1 - \frac{d[ln(d_i d_{i-1})]}{dt} \right\} \right\rangle \tag{8.6}$$

where $x_i = d_i d_{i-1}/\tau$. This is only the direct contribution. The restriction changes the distribution of paths, so all terms in the energy are affected by Fermi statistics. In the above formula, one can neglect the derivative of the nodal distance on "time". This is appropriate at low temperature but could cause some errors at intermediate

temperatures. Note that in computing the energy, all the time slices contribute, not just the reference point, since all the points contribute to the action.

Let us now look at how the restriction is calculated during the path integral Monte Carlo random walk. Consider the case where the trial function is a Slater determinant of single particle density matrices as in Eq. (7.4). Typically in the Monte Carlo random walk only a few (say p) particles are moved at once. This means only p columns of the Slater matrix are changed. Recalculation of the full determinant will take $\mathcal{O}[N^3]$ operations but an update of a p atoms is much more rapid as was shown by Ceperley, Chester and Kalos (1977). The matrix of $g(\mathbf{r}_{it}, \mathbf{r}_{j*}; t) = g_{ijt}$ and its transposed inverse c_{ijt}, defined as the solution to:

$$\sum_k g_{ikt} c_{jkt} = \delta_{ij} \tag{8.7}$$

are computed and saved throughout the run. Using c, we can determine very quickly if a proposed trial move has a positive trial density matrix, because c is also the cofactor matrix of g.

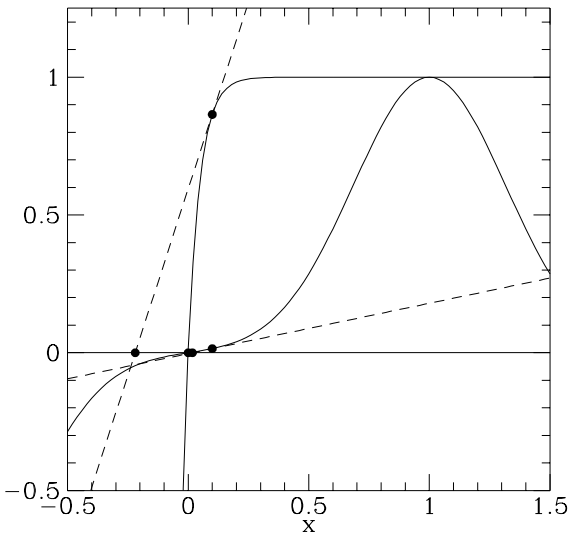

FIG. 8. The antisymmetric density matrix for reference point at 1 and the current point at 1/2. Two ways of estimating the distance to the nodes, both using the Newton-Raphson formula, are shown. In the first method the distance is inversely proportion to the logarithmically derivative of the density matrix, leading to an underestimate of the distance. In the second method, the estimate is the logarithmic derivative of the nearest neighbor is subtracted out. This leads to a distance estimate too large.

Suppose a single particle with label i at time t is moved to a trial position: $\mathbf{r}_{it} \rightarrow \mathbf{r}'$. Then we compute the trial density matrix as:

$$\rho_T(R', R_*; t) = \rho_T(R, R_*; t) \sum_k g(\mathbf{r}', \mathbf{r}_{k*}; t) c_{ikt}. \tag{8.8}$$

This takes $\mathcal{O}[N]$ operations. If the trial density matrix is negative, the trial move is rejected. Otherwise, we go back and recompute the new distances and accept or reject on the resulting changed nodal action. The derivative needed for the distance in the nodal action is determined as:

$$\nabla_i' \ln \rho_T(R', R_*; t) = \sum_k \nabla' g(\mathbf{r}', \mathbf{r}_{k*}; t) c_{ikt} \tag{8.9}$$

Once a move is accepted, we must then update c_{ikt}. The Sherman-Morrison (1949) formula expresses how to correct an inverse that is changed by a single column

$$c'_{jkt} = c_{jkt} + [\delta_{ji} - b_j] c_{ikt} / b_i \tag{8.10}$$

where $b_j = \sum_k g_{ikt} c_{jkt}$ The update takes $\mathcal{O}[N^2]$ operations. After each update, we verify a few entries of Eq. (8.7). If the inverse is no longer accurate, the inverse is recomputed using Gaussian elimination with pivoting. This only happens every several thousand steps, but occurs more often as the number of particles increases or the temperature is lowered since overlapping orbitals make the Slater matrix ill-conditioned. Good update procedures for the more complicated backflow determinants have not been developed. It would be unfortunate if the more accurate trial density matrices were much slower to compute.

B. Sampling restricted paths

The other crucial technical aspect of path integrals is the question of how to sample the path integrals and how to do the permutational sum. While, in principle, one could use the Molecular Dynamics method for the path average, there seems to be no way of performing the permutational sum with MD, so in practice all calculations where particle exchange is important have used a generalization of Metropolis Monte Carlo. In single slice sampling, a single atom in a single time-slice is moved and the new trial position is accepted or rejected based on the action of the new path. However, because the "beads" on the "polymer" are hooked together, uncorrelated motion of individual beads is damped, so that the polymer as a whole moves very slowly. Multiple-slice sampling methods have been devised which speed up the motion through path space. These and the methods for moving through permutation space are discussed at length in Ceperley (1995) for bosonic path integrals.

Since the restricted path method is essentially built on top of the bosonic method, all of those techniques can

be used for fermion path integrals. Here we just remark on a few of the differences. First there is the question of moving non-reference point beads. We move them as if they were bosons, then once an acceptable bosonic path is found, we check the fermion action. First we check whether the new trial path has a positive action at all the new points. If it doesn't that path is immediately rejected. If it is positive, then we compute the estimated distance to the nodal surface, and the image action. This is used to finally accept or reject the path.

The second difference is how to move through permutation space. For bosonic problems, we try inserting cyclic permutations of two, three or four atoms at a time. For fermions we try only three atoms at a time since permutations for two or four atoms would certainly be rejected. A table of the free particle density matrix changes for all possible triplet permutations between particles with the same spin is constructed. This table is used to sample the permutation change. After the permutation is sampled, an appropriate path is sampled.

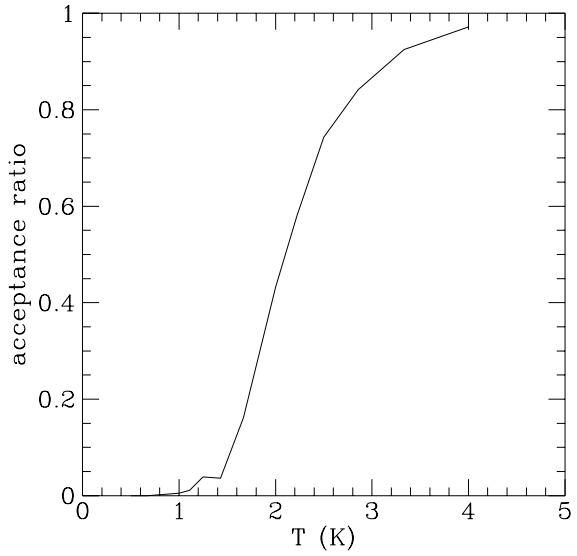

FIG. 9. The acceptance ratio as a function of temperature for liquid helium for a reference point relabeling move. In this move, another time slice is chosen at random to become the reference point. The move is accepted if the current path with the new reference point is node-avoiding. It rapidly approaches zero at low temperatures.

Finally, there is the question of how to move the reference point. If the reference point is moved, the trial density matrix must be recomputed for all $M-1$ links of the path. If any are found to be negative or too close to zero the reference point move is rejected. The action seen by the reference point is non-local. One might worry that this will trap the reference point, and the path as a whole.

In a reference point relabeling move, one does not change the physical location of the paths but merely shifts the whole path in "time". For example one could try to move the reference point from time t_1 to time t_2. The bosonic action is not affected by this shift so it does not have to be recomputed. One simply has to see if the path with the new reference point is node-avoiding. Rejections occur only because the nodes have a dependence on time. In figure 9 is shown the acceptance ratio of this reference relabeling moves in liquid ^{3}He using free-particle nodes. It drops quickly to zero once the system becomes a good Fermi liquid. The reference point strongly resists any uniform shift in the time direction.

We have found that one can treat movement of the particles at the reference point just like the other other time-slices. One moves a few atoms to a new trial position and accepts or rejects that move. In this case we have to check whether that causes the trial density matrix to become negative at some later time. However, if only a single atom is being moved at the reference point., the acceptance ratio does not go to zero at low temperatures.

Trapping has been seen in a related method for the lattice Hubbard model. The constrained path algorithm is similar to restricted path integral Monte Carlo, except the paths are in determinant space, not in configuration space. The algorithms both have a non-local fermion action. Fahy and Hamann (1990) had considered a path integral version of the algorithm and had seen trapping and non-ergodic behavior. In a recent paper, Zhang, Carlson and Gubernatis (1995), showed that the zero-temperature diffusion Monte Carlo algorithm does not have this trapping.

We believe that sampling techniques can significantly reduce the likelihood of trapping but clearly non-ergodic behavior is possible in principle at sufficiently low temperatures or for large numbers of particles. Zero temperature algorithms appear not to suffer from trapping, but the importance sampling which is introduced can bias the calculated result. More experience is needed to find out if trapping is a generic problem for finite temperature path integrals, and whether there are general solutions to this problem.

IX. PERMUTATIONS, THE MOMENTUM DISTRIBUTION AND FERMI LIQUIDS

Here we discuss further qualitative characterization of restricted paths. In our bosonic paper (Ceperley, 1995) was discussed many relationships between paths and physical properties. Much less is known for restricted fermion path integrals but the nodal restriction and the extra spin degrees of freedom make them poten-

tially much richer. Since one can compute the fermion partition function with restricted paths and all thermodynamic properties can be obtained by differentiating the partition function, the paths should reflect any kind of ordering that the fermion system undergoes.

Superfluidity in liquid ^{3}He was first detected by measuring the specific heat. It is interesting to speculate how this triplet pairing will appear with restricted fermions paths. As Feynman(1953) first pointed out, superfluidity in bosonic system is signaled in path integrals by the appearance of permutation cycles of macroscopic size. This has a different interpretation for fermion paths: long permutations are related to the formation of a Fermi liquid, as we now describe.

Let us consider how to calculate the momentum distribution with restricted paths. The momentum distribution is the Fourier transform of the single particle density matrix. For a translationally invariant system:

$$n_{\mathbf{k}} = \frac{1}{8\pi^3} \int d\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} n(\mathbf{r}) \tag{9.1}$$

where the single particle density matrix is:

$$n(\mathbf{r}) = \frac{1}{Z} \int dR \rho(\mathbf{r}_1 + \mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_n, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N; \beta). \tag{9.2}$$

The momentum distribution is the Fourier transform of an off-diagonal element of the density matrix. The paths that we have been discussing up to this point, where all of the paths end at the start of another particle's path, cannot be used to calculate the momentum distribution. To get an observable in momentum space, we cannot do the simulation entirely in the position representation. All that is needed to get the momentum distribution is to allow one of the atoms to have free ends.

However, now the fermion sign comes back in, as it must. First, consider the classical Maxwellian distribution for which:

$$n_{\mathbf{k}} = \left(\frac{\beta\lambda}{\pi}\right)^{3/2} \exp(-\beta\lambda \mathbf{k}^2) \tag{9.3}$$

$$n(\mathbf{r}) = \exp(-\frac{\mathbf{r}^2}{4\beta\lambda}) \tag{9.4}$$

Here the two ends of the cut polymer have an end-to-end vector $\mathbf{r}$ and typically will be within a thermal wavelength of each other. At high temperature, the polymers do not exchange and the shape of an individual atom's path is a pure Gaussian random walk. Hence, both the single particle density matrix and the momentum distribution are Gaussian; both smooth analytic functions.

Now, consider the ideal fermi-gas momentum distribution (for spinless fermions).

$$n_{\mathbf{k}} = \begin{cases} 1/(2\pi^3\rho) & \text{for } k < k_F \\ 0 & \text{for } k > k_F \end{cases} \tag{9.5}$$

$$n(\mathbf{r}) = \frac{3}{(k_F r)^3}[sin(k_F r) - k_F r cos(k_F r)] \tag{9.6}$$

where the Fermi wavevector for spinless fermion is related to the density by $k_F = (6\pi^2\rho)^{1/3}$. Note that the single particle density matrix is proportional to the spherical Bessel function $j_1(z)$, and slowly decays to zero at large r. It has zeroes at $k_F r = 4.493, 7.725, \ldots$. These zeroes mark the places where the even and odd permutations cancel out. Since $n(r)$ is often negative, even with restricted paths we must have negative weights entering. The momentum distribution has a discontinuity at the Fermi wavevector $\mathbf{k}_F$. As a consequence the single particle density matrix must decay at large distance as r^{-2}. We can get such long-range behavior only if there are macroscopic exchanges. Hence the existence of any kind of non-analytic behavior (we mean a discontinuity in $n_{\mathbf{k}}$ or in any of its derivatives) implies that the restricted paths have important macroscopic permutation cycles.

The calculation of the momentum distribution using restricted paths has not yet been performed. The Monte Carlo procedure would include a pair exchange between the cut end and another atom. Remember that for diagonal simulations, only three-body permutations are allowed by the restriction. The pair moves generate the negative contributions to the single particle density matrix.

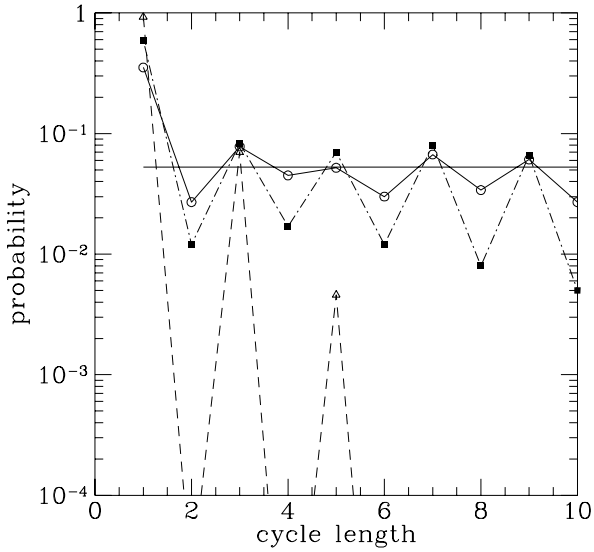

FIG. 10. The probability that an atom is on a permutation with n other atoms shown at three temperatures, the lowest curve with triangles is at 2K, with squares is at 1K, with circles 0.5K. The horizontal line at 1/19 is the expected distribution for a random permutation.

In figure 10 is shown the permutational cycle distribution for liquid ^{3}He, a typical Fermi liquid. It is seen that at temperatures below 0.5K the cycle distribution approaches a uniform distribution. Above this temperature, there is a substantial even-odd effect in the cycle distribution function.

X. OTHER FERMION METHODS

Let me briefly mention several other techniques to handle fermion path integrals in the continuum.

A. Hall's method

A very similar method to what we have described here is an approximate projector method(Hall, 1991, Hall and Price, 1991, and Hall, 1992). Hall argues that walks that are near an exact node of the density matrix end up not contributing to the value of the partition function (or any other property) but do contribute to the variance. Hall discards all paths for which $|\rho_T(R_t, R_\star; \beta)| < \epsilon$ with the method becoming exact as $\epsilon \to 0$. Note that it is possible to get negative contributions; the walks can "hop" over the nodes. Hall argues that if $\epsilon >> \tau$ one gets rid of most of the negative signs, but the result is approximate. If $\epsilon << \tau$, hopping is easy, the method is exact but the variance is just as large as it was before the restriction. Hence, ϵ and τ are parameters that can be used to control independently the statistical error and the systematic error.

From the restricted-path point-of-view we can understand how Hall's method can work. If we eliminate paths close to a node, we eliminate equal numbers of positive and negative paths and this will not change the answer. But by formulating the restriction as a boundary condition instead of as a condition on the value of the density matrix, we can make better nodal actions, and work at larger values of ϵ and τ without losing accuracy.

Hall also symmetrizes the action with respect to which point is considered as the reference point. This will increase the computational time since M^2 checks of the trial density matrix have to be made versus M if you do not symmetrize. Our experience with liquid ^{3}He for temperatures below the degeneracy temperature, using free-particle trial density matrix, shows that only a single point at a time can be the reference point for a typical path. Symmetrizing with respect to the reference point only increases the computational work, without bring any increased efficiency.

Hall's method may be useful in making a path integral version of the release-node method that is used in ground-state Quantum Monte Carlo to go beyond the fixed-node approximation. One would like to have a method to add in a few node-crossing walks to fix up the nodal surfaces locally, but in a controlled fashion so that the efficiency is no more than one order-of-magnitude lower than the restricted-path efficiency.

B. Slice-wise antisymmetrization

One is allowed to apply the antisymmetrization projection operator on each time slice as suggested by Takahashi and Imada (1984). The action contains a determinant at each slice, not a restriction. The product of the signs of the determinants is used to weight the paths as discussed in Sec. (3). In contrast to the restricted path integral approach, one retains time slice symmetry. In one spatial dimension this solves the sign problem since one picks up a sign when two fermions hop over each other, but if one antisymmetrizes each step, the positive sign path always be greater than the negative path. Consider this procedure as applied to the hydrogen molecule in Sec. (5). The probability of a path will be the bosonic action times a factor:

$$\rho_F \propto \prod_{k=1}^{M} [1 - \exp(-\mathbf{r}_k \cdot \mathbf{r}_{k-1}/(\lambda\tau))] \qquad (10.1)$$

where $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. This prevents an exchange in a single step (actually it only rules out steps with $|\mathbf{r}_k \cdot \mathbf{r}_{k+1}| \leq \lambda\tau$). But it does not rule out paths of type B although it may diminish their probability somewhat if the thermal wavelength is comparable to the bond-length. The antisymmetrization has a spatial range $r \approx (\lambda\tau)^{1/2}$ so it is only a local modification of the action. At low temperature there will still be a cancellation between paths A and C against those of B resulting in an exponentially bad sign problem.

Newman and Kuki(1992) have developed this technique further by averaging also over other rotational symmetries. They achieved much higher efficiencies for two electron problems than with the direct method. They still observed an exponentially decreasing efficiency, as β increased, but the rate of growth was much diminished. With increasing numbers of particles the effect of this symmetrization is less. If the boson system can undergo a lambda transition, the fermion efficiency is guaranteed to be exponentially small. The conclusion is that single-slice antisymmetrization will not help much for interesting fermion systems. Also, it does not seem easy to combine it with the restricted path integral approach. One has to choose between antisymmetrizing at every slice, or restricting the paths to be node-avoiding.

C. Ignoring the sign

Suppose one does as much slicewise antisymmetrization as possible but ignore the sign of the negative walks. This was suggested by Imada (1984), giving the following ingenious statistical argument:

> In principle, even one sample [of an integrand] can give a good convergence to the exact value if the calculation is performed in the thermal equilibrium state of a sufficiently large system. It should be noted that the thermal equilibrium state is realized in a sample generated by the importance sampling method, in which the configurations are generated in proportion to the contribution to the partition function. If we regard the generated samples of a finite system as a sequence in time evolution, our assertion seems to be related to the ergodic hypothesis: A long time average is equal to the phase average. If we calculate the average from only one sample Eq. (3.2) is replaced by
>
> $$\langle X \rangle \approx X_1 \approx X_2 \approx X_3 \ldots \approx X_G. \qquad (10.2)$$
>
> If the system size is infinite, X_i should not depend on i. When the system size is large but finite in the actual simulation, better convergence is expected from the average:
>
> $$\langle X \rangle \approx \sum_{i=1}^{G} X_i/G. \qquad (10.3)$$
>
> The finite size effect in Eq. (10.3) must decrease with increasing system size. Equation (10.3) is a surprisingly simple result. This can be calculated without concern regarding the sign of the weight function. As similar to the ergodicity in classical systems, the validity of Eq. (10.3) in sufficiently large systems seems to be obvious from a physical point of view, although it has not been mathematically proven yet.

He concludes that ignoring the sign is an exact procedure for a sufficiently large system.

Such a procedure cannot be correct. As we have explained above, when one ignores the sign, one is using an effective bosonic action. If the system remains liquid as it is cooled, it must Bose condense. One will get a completely different behavior in a fermion system. The above reasoning is flawed because classical statistical mechanics applies to a probability distribution, not to an integrand which can be positive and negative. [See also Sorrella et al (1988,1989) for a related method for lattice models and the analysis of the Sorrella method by Loh et al (1990).]

D. Cancellation

Cancellation schemes have been developed for zero temperature quantum Monte Carlo, see for example Anderson (1995). By cancellation, we mean that one starts with a population of positive and negative random walks. If a positive and negative random walk approach each other, they annihilate. Alternatively, they could repel each other. This formally solves the sign problem. The remaining sticky point is to determine how many walks are needed to create a stable interface between the positive and negative walks. Estimates show that the minimum stable population size will grow exponentially in the number of fermions. Hence, this has not yet led to a viable many-fermion algorithm. Such methods have not yet been developed for finite-temperature path integrals.

XI. CURRENT APPLICATIONS OF FERMION PATH INTEGRALS

I will not make a detailed comparison between the path integral approach for quantum simulations with the other methods such as Variational Monte Carlo, projector Monte Carlo (Greens function or Diffusion) or ab initio molecular dynamics using forces from LDA. There is a comparison between boson path integrals and zero temperature quantum Monte Carlo methods in Ceperley (1995). Path Integrals are unique in being able to treat strongly correlated systems, and mixed quantum-classical systems, including the effects of temperature, and having other, more technical advantages.

I also think that they are unique in being amenable to a "black-box" approach since no trial function is used, only a trial restriction. The best proof of this assertion is that applications of RPIMC to a variety of systems are already beginning to appear. Some of the applications are considerably more complicated that have been treated with zero-temperature quantum Monte Carlo methods.

The first application of RPIMC was to liquid ^{3}He by Ceperley (1992) using the semi-empirical Aziz potential. Using free-particles nodes, these calculations showed reasonable agreement with experiment and also showed the importance of using time-dependent nodes and permutations. See figure 11. There has also been a recent study of the interface between superfluid ^{4}He and liquid ^{3}He by Boninsegni and Ceperley (1995) who were able to determine the width of the interface and estimate directly solubilities and effective masses in the ^{3}He-rich and ^{4}He rich phase.

In our group, there has been a major effort to calculate properties of an electron-proton hydrogenic plasma by Pierleoni *et al.* (1994) and Magro *et al.* (1994). In these calculations, from 32 to 64 electrons and an equal

number of protons were put into a periodic box, interacting only with the Coulomb potential. Both the electrons and protons were fully quantum particles. A time step of roughly 10^6K was found adequate to give accurate results. We were thus able to reach temperatures of 4000K using 250 time slices. Using reasonable length runs on CRAY-YMP computers and fast workstations we found good agreement with other theoretical approaches at high density and at high temperature in the plasma phase.

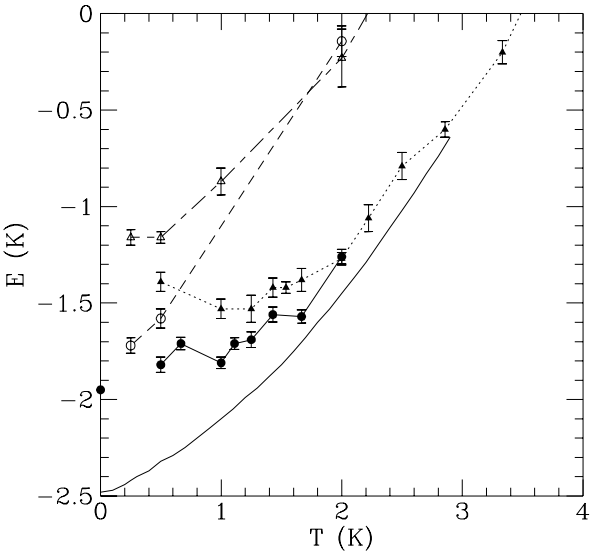

FIG. 11. The energy of liquid ^{3}He vs temperature at a density of 0.16355 Å^{-3} (near zero pressure). The line represents the experimental data. The open symbols are the RPIMC calculations using ground state nodes; the filled symbols are RPIMC calculations with the free-particle density matrix nodes. The circles are with permutations, the triangles are with the identity permutation. The zero temperature result was obtained with GFMC with free-particle nodes.

At temperatures below 10,000K we observed the spontaneous formation of molecules, as evidenced by a strong peak in the proton-proton correlation function at a distance of 1Å. At slightly lower temperature of 7,000K we observed a very rapid build up in the molecular density, enough to suggest that there is a first-order phase transition from the plasma state into a molecular liquid. The existence of the so called "plasma phase transition" has been guessed at for a few years. It can be crucial for understanding the interior of the giant planets or, more generally, the phase diagram of hydrogen at high temperatures and pressures. The ability to study such phase transitions at finite temperature shows the power of the new method. Also striking is the lack of theoretical input to the calculation. Equivalent Variational and Greens Function Monte Carlo calculations at zero temperature (Ceperley and Alder, 1987, Natoli ***et al***, 1993, 1995) required careful construction and optimization of trial wave functions.

These various simulations demonstrate the power of the method. Once the RPIMC method is programmed, one can rather directly do highly accurate simulations of experimentally relevant systems without tedious construction of basis functions and trial functions; the type of systems which can be treated are much more complex than can be handled with other techniques.

XII. FUTURE PROSPECTS

We think that the Path Integral Monte Carlo method is a very powerful method capable of doing very accurate simulations of both boson and fermion systems. Its methodology is much less well developed than for classical systems and there are many challenges in applying this method to real systems. Among them are:

1. Clearly much work needs to be done in figuring out what we should use as nodes since the restriction is the only uncontrolled approximation. Free-particle nodes are too primitive and possibly pathological in that they do not have any coupling from unlike spins. Analytical and numerical methods are needed to go from the free particle nodes at high temperature to the mean-field-like nodes that are useful at low temperature for atoms and molecules. We are currently working on this. We need to clarify the situation with regard to free-energy bounds so as to internally (within the RPIMC method) determine optimal nodes.

2. We need ways to get to lower temperatures. One of the advantages of the RPI approach is that systems with both quantum and classical degrees of freedom are treated more easily than in other approaches. But to exploit this possibility, we need to simulate electrons at room temperature. So far, our lowest temperature with many-body hydrogen including both electron and proton paths has been 3,000K. There are several ways of going to lower temperature, for example one could use zero-temperature nodal surfaces for the electrons. But much experience is needed to verify these approximations.

3. One of the big advances in recent years in zero-temperature quantum Monte Carlo methods has been in the development of pseudo-potential methods. This has allowed one to treat systems with carbon, nitrogen, silicon atoms as easily as hydrogen and those with transition metal atoms (*e. g.* iron and copper) with an order of magnitude more effort. Two basic approaches have been

developed. In the pseudo-Hamiltonian approach (Bachelet, Ceperley and Chiocchetti, 1989), one alters the electron mass inside the core so one can use directly all the quantum Monte Carlo methods. With this approach, one cannot treat all types of atoms sufficiently accurately. In the more general, non-local pseudo-potential method one has a different potential for each angular momentum state of the electron. The angular momentum dependence causes problems with the nodal restriction since paths are not continuous. What has been done (Mitas et al 1991) is to project the non-locality onto an accurate trial function and thereby make it local. One only requires accurate trial functions in the core region. Perhaps the use of atomic wave functions will be sufficient to make this work in path integrals. We do not want to have to bring many-body trial functions back in.

4. Path Integral Monte Carlo methods are computationally expensive. This has limited the number of applications. The easy answer to this objection is that computers are getting faster and cheaper. Nonetheless, improving the efficiency of the methods will allow us to solve much more difficult or realistic problems. More effort needs to go into ways of speeding up the dynamics of moving the paths, particularly at low temperature. We also need better nodal actions so the time step can be made larger. Finally, we need to write $\mathcal{O}$[N] codes so we can do much larger systems. (Note that RPIMC is naturally $\mathcal{O}$[N] since the single-electron density matrix in Eq. (7.5) is localized in the range of the thermal wavelength.)

5. The computation of dynamical properties is a challenge. As Gubernatis has described there is much progress with maximum entropy methods. RPIMC has gotten rid of the sign problem in calculating the partition function but what is needed for dynamics are off-diagonal density matrix elements. We have gotten rid of most but not all of the negative permutations on the diagonal (Boninsegni and Ceperley 1994). Because of these signs, it seems unlikely that RPIMC will be able to calculate the highly accurate imaginary-time correlation functions needed for maximum entropy reconstruction of the real-time dynamics. As a consequence, properties such as conductivity, and spectral functions, seem very difficult to compute with these methods. Perhaps one needs to formulate dynamical properties in terms of thermodynamic properties as done with the superfluid density of bosonic systems.

6. Much needs to be done in understanding how to interpret RPI. For example, what is the description of superfluidity, the Mott transition, molecular bonding and bound versus continuum states? Girardeau (1990) has suggested that eigenvalues and eigenfunctions of the one- and two-body density matrices can be used to identify the number of molecules. Such eigenfunctions (the natural orbitals) are computable with path integrals as we described above with the momentum distribution. This shows how to make the connection between properties of the paths and rigorous many-body theory. There are a few hints that many of the relationships we have discovered for bosonic systems have equally good interpretations in terms of RPI.

7. The generalization of the method to magnetic fields is possible, see Ortiz et al (1993). It will also be interesting to work with other ensembles, for example with fixed angular momentum to see vortices.

The above is a list of challenges for those who wish to develop further fermion path integrals. Hopefully, we will have some answers by the next simulation meeting at Lake Como.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research (N00014-90-J-1783), by the National Center for Supercomputing Applications and the Department of Physics at the University of Illinois Urbana-Champaign. The computations were done using the facilities at NCSA. I would like to acknowledge useful discussion with my colleagues: W. R. Magro, M. Boninsegni, B. Bernu and C. Pierleoni and to the school's organizers for inviting me to present and write up this contribution. More recent developments may be found at http://www.ncsa.uiuc.edu/Apps/CMP. I welcome correspondence and corrections: ceperley@uiuc.edu.

Anderson, J. B., 1976, J. Chem. Phys. **65**, 4121.

Anderson, J. B. Int. Rev. Phys. Chem, **14**, 85; and *Understanding Chemical Reactivity*, ed. S. R. Langhoff, Kluwer, Dordrecht.

Bachelet, G.B., D. M. Ceperley, and M. Chiocchetti, 1989, Phys. Rev. Letts., **62**, 208.

Boninsegni, M., and D. M. Ceperley, 1994, unpublished.

Boninsegni, M., and D. M. Ceperley, 1995, Phys. Rev. Lett. **74**, 2288.

Buch, V., 1994, J. Chem. Phys. **100**, 7610.

Burkard, R. E., and U. Derigs, 1980, *Assignment and Matching Problems*, (Springer-Verlag, Berlin).

Ceperley, D. M., 1991, J. Stat. Phys. **63**, 1237.
Ceperley, D. M., 1992, Phys. Rev. Lett. **69**, 331.
Ceperley, D. M., 1995, Rev. Mod. Phys. **67**, 279.
Ceperley, D. M. and B. J. Alder, 1980, Phys. Rev. Lett. **45**, 566.
Ceperley, D. M. and B. J. Alder, 1987, Phys. Rev. **B 36**, 2092.
Ceperley, D., Chester, G.V., and Kalos, M.H., 1977, Phys. Rev. **B 16**, 3081.
Fahy, S. B. and D. R. Hamann, 1990, Phys. Rev. Lett. **65**, 3437.
Feynman, R. P., 1953, Phys. Rev. **90**, 1116; Phys. Rev. **91**, 1291; Phys. Rev. **91**, 1301.
Feynman, R. P., and A. R. Hibbs, 1965, *Quantum Mechanics and Path Integrals*, (McGraw-Hill, New York).
Girardeau, M. D. 1990, Phys. Rev. **A41**, 6935.
Hall, R. W., 1991 J. Chem. Phys. **94**, 1312.
Hall, R. W. 1992 J. Chem. Phys. **97**, 6481.
Hall, R. W. and M. R. Price, 1991 J. Chem. Phys. **95**, 5999.
Hammond, B. L., W. A. Lester, Jr., and P. J. Reynolds, *Monte Carlo Methods in ab initio Quantum Chemistry*, World Scientific, Singapore, 1994.
Imada, M. 1984, J.Phys. Soc. Japan, **53**, 2861.
Kwon, Y., D. M. Ceperley and R. M. Martin, 1993, Phys. Rev. **B48**, 12037.
Landau, L. D. and E. M. Lifshitz, 1970, *Statistical Physics*, Addison-Wesley.
Loh E.Y., J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar, 1990, Phys Rev **B41**, 9301.
Lyubartsev, A. P., and P. N. Vorontsov-Velyaminov, 1993 Phys. Rev. **A48**, 4075.
Magro, W. 1994, University of Illinois Ph. D. Thesis.
Mitas, L., Shirley E. L., and Ceperley, D. M., 1991, J. Chem. Phys. **95**, 346.
Natoli, V., R. M. Martin and D. Ceperley, 1993, Phys. Rev. Letts. **70**, 1952; 1995, Phys. Rev. Letts. 1601.
Newman, W. H. and Kuki, A., 1992, J. Chem. Phys. **96** 1409.
Ortiz, G., Ceperley, D. M. and Martin R. M.,1993, Phys. Rev. Letts. **71**, 2777.
Pierleoni, C., B. Bernu, D. M. Ceperley, and W. R. Magro, 1994, Phys. Rev. Lett. **73**, 2145.
Panoff, R. M. and J. Carlson, 1989, Phys. Rev. Letts. **62**, 1130.
K. E. Schmidt and M. H. Kalos 1984, in *Monte Carlo Methods in Statistical Physics II*, ed. K. Binder, Springer, 1984.
Runge, K. J., M. P. Surh, C. Mailhiot and E. L. Pollock, 1992, Phys. Rev. Letts. **69**, 3527.
Senatore, G. and N. H. March, 1994, Rev. Mod. Phys. **66**, 445.
Sherman, J., and W. J. Morrison, 1949, Ann. Math. Statist., **20**, 621.
Sorrella, S., S. Baroni, R. Car, and M. Parrinello, 1989, Europhys. Letts. **8**, 663.
Sorrella, S., E. Tosati, S. Baroni, R. Car, and M. Parrinello, 1988 Int. J. Mod. Phys., **B1**, 993.
Takahashi, M. and M. Imada, 1984, J. Phys. Soc. Japan, **53**, 963.
Zhang, S. , J. Carlson, and J. E. Gubernatis, 1995, Phys. Rev. Letts., **74**, 3652.